Estimating multivariate GARCH and stochastic correlation models equation by equation

Francq, Christian and Zakoian, Jean-Michel

2014

Estimating multivariate GARCH and stochastic correlation models equation by equation

Christian Francq

CREST and Université Lille 3 (EQUIPPE)

Jean-Michel Zakoïan

CREST and Université Lille 3 (EQUIPPE)

March 8, 2014

Abstract. A new approach is proposed to estimate a large class of multivariate volatility models. The method is based on estimating equation-by-equation the volatility parameters of the individual returns by quasi-maximum likelihood in a first step, and estimating the correlations based on volatility-standardized returns in a second step. Instead of estimating a d-multivariate volatility model we thus estimate d univariate GARCH-type equations plus a correlation matrix, which is generally much simpler and numerically efficient. The strong consistency and asymptotic normality of the first-step estimator is established in a very general framework. For generalized constant conditional correlation models, and also for some time-varying conditional correlation models, we obtain the asymptotic properties of the two-step estimator. Our estimator can also be used to test the restrictions imposed by a particular MGARCH specification. An application to financial series illustrates the interest of the approach.

Keywords: Constant conditional correlation, Dynamic conditional correlation, Markov switching models, Multivariate GARCH, Quasi maximum likelihood estimation.

Address for correspondence: Jean-Michel Zakoïan, CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff cedex.

Email: zakoian@ensae.fr

1. Introduction

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models have featured prominently in the analysis of financial time series. The versions initially stressed in the econometric literature (see Engle (1982) and Bollerslev (1986)) are univariate. The last twenty years have witnessed significant research devoted to the multivariate extension of the concepts and models developed for univariate GARCH. Among the numerous specifications of multivariate GARCH (MGARCH) models, the most popular seem to be the Constant Conditional Correlations (CCC) model introduced by Bollerslev (1990) and extended by Jeantheau (1998), the BEKK model developed by Baba, Engle, Kraft and Kroner, in a preliminary version of Engle and Kroner (1995), and the Dynamic Conditional Correlations (DCC) models proposed by Tse and Tsui (2002) and Engle (2002). Reviews on the rapidly changing literature on MGARCH are Bauwens, Laurent and Rombouts (2006), Silvennoinen and Teräsvirta (2009), Francq and Zakoïan (2010, Chapter 11), Bauwens, Hafner and Laurent (2012).

The complexity of MGARCH models has been a major obstacle to their use in applied works. Indeed, in asset pricing applications or portfolio management, cross-sections of hundreds of stocks are common. However, as the dimension of the cross section increases, the number of parameters can become very large in MGARCH models, making estimation increasingly cumbersome. This "dimensionality curse" is general in multivariate time series, but is particularly problematic in GARCH models. The reason is that the parameters of interest are involved in the conditional variance matrix, which has to be inverted in gaussian likelihood-based estimation methods.

Existing approaches to alleviate the dimensionality curse rely on either constraining the structure of the model in order to reduce the number of parameters, or using an alternative estimation criterion. Examples of models belonging to the first category are the Factor ARCH models of Engle, Ng and Rotschild (1990), the Generalized Orthogonal GARCH model of van der Weide (2002), and the Generalized Orthogonal Factor GARCH model of Lanne and Saikkonen (2007). The second strategy was advocated by Engle, Shephard and Sheppard (2008), who suggested to use a composite likelihood instead of the usual quasi-likelihood. An approach combining the two concepts, reduction of the parameter dimension and use of a partial likelihood, was recently proposed by Engle and Kelly (2012) who introduced the Dynamic Equicorrelation (DECO) model.

A solution to the high-dimension problem which does not preclude a high-dimensional parameter set relies on estimating *equation by equation* the conditional variances of each component of a vector of returns. The conditional variance of component k is a function, parameterized by some parameter vector $\boldsymbol{\theta}^{(k)}$, of the past of *all components* of the vector of returns. Thus, the univariate models of the components are generally not GARCH in the classical sense. This approach has been used in several empirical studies (see e.g. Sucarrat, Grønneberg and Escribano (2013) for a recent reference) but asymptotic results are lacking. We propose an Equation-by-Equation (EbE) method of estimation, based on the Quasi-Maximum Likelihood (QML), for the volatility parameters $\boldsymbol{\theta}^{(k)}$ and, under appropriate assumptions, we develop an asymptotic theory for such EbE estimators (EbEE).

Apart from the numerical simplicity, one advantage of this approach is that the derivation of EbEE is independent from the specification of a conditional correlation matrix. It can therefore be employed for CCC GARCH models as well as for DCC GARCH models, leading to the same estimators of the individual volatilities. It can also be used for multivariate models that are not GARCH. We consider a class of *Stochastic Correlation* (SC) models which has the same multiplicative form as GARCH-type models, except that the correlation matrix is not a measurable function of the past observations. The term stochastic correlation obviously refers to the class of Stochastic Volatility models, which differ from GARCH by the fact that the volatility depends on unobservable stochastic factors.

Estimation of the individual conditional variances can be completed, in a second step, by the estimation of a time-varying correlation matrix using the standardized returns obtained in the first step. For CCC models, the constant conditional correlation matrix can be estimated by the empirical correlation matrix of the EbEE residuals. For some DCC and SC models, the structure of the time-varying correlation can also be estimated by modeling the dynamics of the EbEE residuals. In this article, we derive asymptotic results for this estimator, which can be seen as an extension of the two-step estimator proposed by Engle and Sheppard (2001) in the case where the individual volatilities have pure GARCH forms with iid innovations. The present paper considers augmented GARCH individual volatilities depending on lagged values of all the components of the returns, with the possibility of volatility spillovers, and also enables the estimation of more complex correlation matrices.

The paper is organized as follows. Section 2 presents the assumptions and notations for the class of multivariate processes studied in this article. Such assumptions are discussed under different assumptions on the correlation matrix $\boldsymbol{R}_t$. In Section 3, we study the estimation of the volatility parameters without any assumption on $\boldsymbol{R}_t$. Section 4 develops the two-step estimation method in the case of constant conditional correlation and stochastic correlation matrices. Consistency and asymptotic normality of the estimator are established. Section 5 studies testing the adequacy of a class of multivariate models. In Section 6, we apply our method to a large set of stock market indices, and to several exchange rate series. Section 7 concludes. The most technical assumptions and the proofs of the main theorems are collected in the Appendix.

2. Models and assumptions

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \cdots, \epsilon_{mt})'$ be a $\mathbb{R}^m$-valued process and let $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$ be the σ-field generated by $\{\boldsymbol{\epsilon}_u, u < t\}$. Assume

$$E(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \mathbf{0}, \tag{2.1}$$

and

$$\mathrm{Var}(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{H}_t \quad \text{exists and is positive definite.} \tag{2.2}$$

Denoting by σ_{kt}^2 the diagonal elements of $\boldsymbol{H}_t$, that is the variances of the components of $\boldsymbol{\epsilon}_t$ conditional on $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$, we introduce the vector

$$\boldsymbol{\eta}_t^* = \boldsymbol{D}_t^{-1}\boldsymbol{\epsilon}_t = (\epsilon_{1t}/\sigma_{1t}, \dots, \epsilon_{mt}/\sigma_{mt})' \quad \text{where } \boldsymbol{D}_t = \mathrm{diag}(\sigma_{1t}, \dots, \sigma_{mt}).$$

By (2.1)-(2.2), we have $E(\boldsymbol{\eta}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \mathbf{0}$ and the conditional correlation matrix of $\boldsymbol{\epsilon}_t$ is given by

$$\boldsymbol{R}_t = \mathrm{Var}(\boldsymbol{\eta}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{D}_t^{-1}\boldsymbol{H}_t\boldsymbol{D}_t^{-1}.$$

It follows that, for $k = 1, \dots, m$,

$$E(\eta_{kt}^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = 0, \qquad \mathrm{Var}(\eta_{kt}^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = 1. \tag{2.3}$$

Introducing the vector $\boldsymbol{\eta}_t$ such that $\boldsymbol{\eta}_t^* = \boldsymbol{R}_t^{1/2}\boldsymbol{\eta}_t$, the previous equations can be summarized as follows. The square root has to be understood in the sense of the Cholesky factorization, that is, $\boldsymbol{R}_t^{1/2}(\boldsymbol{R}_t^{1/2})' = \boldsymbol{R}_t$ and $\boldsymbol{H}_t^{1/2}(\boldsymbol{H}_t^{1/2})' = \boldsymbol{H}_t$.

ASSUMPTIONS AND NOTATIONS: *The $\mathbb{R}^m$-valued process $(\boldsymbol{\epsilon}_t)$ satisfies*

$$\left\{\begin{array}{rcl}\boldsymbol{\epsilon}_t &=& \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \qquad E(\boldsymbol{\eta}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \mathbf{0}, \quad Var(\boldsymbol{\eta}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{I}_m, \\ \\ \boldsymbol{H}_t &=& \boldsymbol{H}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots) = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t,\end{array}\right. \tag{2.4}$$

where $\boldsymbol{D}_t = diag(\boldsymbol{H}_t^{1/2})$ *and* $\boldsymbol{R}_t = Corr(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}})$.

We assume that the conditional variance of the k-th component of $\boldsymbol{\epsilon}_t$ is parameterized by some parameter $\boldsymbol{\theta}_0^{(k)} \in \mathbb{R}^{d_k}$, so that

$$\left\{\begin{array}{rcl}\epsilon_{kt} &=& \sigma_{kt}\eta_{kt}^*, \\ \sigma_{kt} &=& \sigma_k(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots; \boldsymbol{\theta}_0^{(k)}),\end{array}\right. \tag{2.5}$$

where $\sigma_k : \mathbb{R}^\infty \times \boldsymbol{\Theta}_k \to (0, \infty)$. In view of (2.3), the process $(\boldsymbol{\eta}_t^*)$ can be called the vector of *Equation-by-Equation* (EbE) innovations of $(\boldsymbol{\epsilon}_t)$.

REMARK 2.1. In this model, the volatility of any component of $\boldsymbol{\epsilon}_t$ is allowed to depend on the past values of all components. For this reason, Model (2.5) can be referred to as an *augmented GARCH* model. Moreover, the innovations η_{kt}^* are not iid. Thus, (2.5) is not a Data Generating Process (DGP).

We now consider two classes of DGP satisfying the previous assumptions.

2.1. GARCH-type models

Consider a GARCH process, defined as a non anticipative[1] solution of

$$\boldsymbol{\epsilon}_t = \boldsymbol{D}_t\boldsymbol{R}_t^{1/2}\boldsymbol{\eta}_t, \quad \text{where } (\boldsymbol{\eta}_t) \text{ is an iid sequence.} \tag{2.6}$$

Obviously, $(\boldsymbol{\epsilon}_t)$ thus satisfies (2.4). A variety of parametric forms of function $\boldsymbol{H}$ has been introduced in the literature. In the GARCH literature, it is usual to distinguish Constant Conditional Correlation (CCC) models, for which

$$\boldsymbol{R}_t = \boldsymbol{R} \text{ is a constant correlation matrix,} \tag{2.7}$$

from Dynamic Conditional Correlation (DCC) models where $\boldsymbol{R}_t$ is a non constant function of the past of $\boldsymbol{\epsilon}_t$, that is,

$$\boldsymbol{R}_t = \boldsymbol{R}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots) \neq \boldsymbol{R}.$$

[1]that is $\boldsymbol{\epsilon}_t \in \mathcal{F}_t^{\boldsymbol{\eta}}$, the σ-field generated by $\{\boldsymbol{\eta}_u, u \leq t\}$.

Note that in the case of CCC models, the sequence $(\boldsymbol{\eta}_t^*)$ is iid which is generally not the case for DCC models.

2.2. Stochastic Correlation Models

To obtain a DGP satisfying (2.4), an alternative to GARCH-type models is to introduce correlation matrices that are not function of the past but also depend on some latent process (Δ_t). More precisely, let

$$\boldsymbol{\epsilon}_t = \boldsymbol{D}_t \boldsymbol{R}_t^{*1/2} \boldsymbol{\xi}_t, \tag{2.8}$$

where $(\boldsymbol{\xi}_t)$ is an iid $(\mathbf{0}, \boldsymbol{I}_m)$ sequence and

$$\boldsymbol{R}_t^* = \boldsymbol{R}^*(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \Delta_t), \qquad \Delta_t \notin \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}. \tag{2.9}$$

By analogy with the so-called Stochastic Volatility models, in which the volatility is not a measurable function of the past observables, we can call model (2.8)-(2.9) a *Stochastic Correlation* (SC) model. For this model, the individual volatilities σ_{kt}, as given by (2.5), are of GARCH-type, while the correlations between components in $\boldsymbol{R}_t^*$ are not. In this context, a non anticipative solution of the model is such that $\boldsymbol{\epsilon}_t \in \mathcal{F}_t^{\boldsymbol{\xi},\Delta}$, the σ-field generated by $\{\boldsymbol{\xi}_u, \Delta_u, u \leq t\}$. Assuming that

$$(\boldsymbol{\epsilon}_t) \text{ is a non anticipative solution and } \boldsymbol{\xi}_t \text{ is independent from } \mathcal{F}_t^{\Delta}, \tag{2.10}$$

the σ-field generated by $\{\Delta_u, u \leq t\}$, we have $E(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \mathbf{0}$, and

$$\boldsymbol{H}_t = \mathrm{Var}(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{D}_t E(\boldsymbol{R}_t^{*1/2} \boldsymbol{\xi}_t \boldsymbol{\xi}_t' \boldsymbol{R}_t^{*'1/2} \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) \boldsymbol{D}_t = \boldsymbol{D}_t E(\boldsymbol{R}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) \boldsymbol{D}_t,$$

using the fact that $E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t') = \boldsymbol{I}_m$. Note that the conditional correlation matrix is $\boldsymbol{R}_t = E(\boldsymbol{R}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}})$.

Therefore, SC models (2.8)-(2.10), which are extensions of GARCH-type models, satisfy Assumptions (2.4). Note that the three innovations sequences are linked by

$$\boldsymbol{\eta}_t^* = \boldsymbol{R}_t^{*1/2} \boldsymbol{\xi}_t = \boldsymbol{R}_t^{1/2} \boldsymbol{\eta}_t.$$

3. Equation-by-equation estimation of volatility parameters

In this section, we are interested in estimating the conditional variance of each component of $\boldsymbol{\epsilon}_t$ satisfying (2.4). In other words, we study the estimation of the parameter $\boldsymbol{\theta}_0^{(k)}$ in Model (2.5), under (2.3), for $k = 1, \dots, m$.

To estimate $\boldsymbol{\theta}_0^{(k)}$ we will use the Gaussian QML, which is the most widely used estimation method for univariate GARCH models, but other methods could be considered as well (for instance the LAD method or the weighted QML studied by Ling (2007), the non Gaussian QML studied by Berkes and Horváth (2004)). In view of Remark 2.1, Model (2.5) is not, in general, a univariate GARCH and we cannot directly rely on existing results for its estimation.

Given observations $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$, and arbitrary initial values $\tilde{\boldsymbol{\epsilon}}_i$ for $i \leq 0$, we define $\tilde{\sigma}_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma_k(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \boldsymbol{\epsilon}_1, \tilde{\boldsymbol{\epsilon}}_0, \tilde{\boldsymbol{\epsilon}}_{-1}, \dots; \boldsymbol{\theta}^{(k)})$ for $k = 1, \dots, m$ and $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}_k$, assuming that $\boldsymbol{\Theta}_k$ is a compact parameter set and $\boldsymbol{\theta}_0^{(k)} \in \boldsymbol{\Theta}_k$. This random variable will be used as a proxy of $\sigma_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma_k(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_{-1}, \dots; \boldsymbol{\theta}^{(k)})$.

Let $\hat{\boldsymbol{\theta}}_n^{(k)}$ denote the equation-by-equation estimator (EbEE) of $\boldsymbol{\theta}_0^{(k)}$:

$$\hat{\boldsymbol{\theta}}_n^{(k)} = \arg \min_{\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}), \quad \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n} \sum_{t=1}^{n} \log \tilde{\sigma}_{kt}^2 \left(\boldsymbol{\theta}^{(k)}\right) + \frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2 \left(\boldsymbol{\theta}^{(k)}\right)}.$$

Similarly, define

$$Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n} \sum_{t=1}^{n} \log \sigma_{kt}^2 \left(\boldsymbol{\theta}^{(k)}\right) + \frac{\epsilon_{kt}^2}{\sigma_{kt}^2 \left(\boldsymbol{\theta}^{(k)}\right)} := \frac{1}{n} \sum_{t=1}^{n} \ell_{kt}(\boldsymbol{\theta}^{(k)}).$$

3.1. Consistency and asymptotic normality of the EbEE

We make the following assumption on the process $(\boldsymbol{\epsilon}_t)$.

A1: $(\boldsymbol{\epsilon}_t)$ is a strictly stationary and ergodic process satisfying (2.4), with $E|\epsilon_{kt}|^s < \infty$ for some $s > 0$. Moreover, $E \log \sigma_{kt}^2 < \infty$.

This assumption can be made more explicit for specific models (see for instance Theorem 2.1 and Corollary 2.2 in Francq and Zakoian (2012)). Technical assumptions on the function σ_k are relegated to Appendix A. Assumptions **A4-A6** are required for the consistency. To prove the asymptotic normality, we need to assume

A7: $\boldsymbol{\theta}_0^{(k)}$ belongs to the interior of $\boldsymbol{\Theta}^{(k)}$,

A8: $E\left|\eta_{kt}^*\right|^{4(1+\delta)} < \infty$, for some $\delta > 0$,

and some additional technical assumptions **A9-A12**.

THEOREM 3.1. *If* **A1** *and* **A4**-**A6** *hold, then*

$$\hat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}, \quad a.s. \quad as\ n \to \infty.$$

If, in addition, **A7**-**A12** *hold, then*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right) \quad \xrightarrow{\mathcal{L}} \quad \mathcal{N}\left\{0, \boldsymbol{J}_{kk}^{-1}\boldsymbol{I}_{kk}\boldsymbol{J}_{kk}^{-1}\right\},$$

where

$$\boldsymbol{I}_{kk} = E\left(\{\eta_{kt}^{*4} - 1\}\boldsymbol{d}_{kt}\boldsymbol{d}_{kt}'\right), \quad \boldsymbol{J}_{kk} = E\left(\boldsymbol{d}_{kt}\boldsymbol{d}_{kt}'\right), \quad \boldsymbol{d}_{kt} = \frac{1}{\sigma_{kt}^2}\frac{\partial\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}^{(k)}}.$$

Note that the sequence of $(\boldsymbol{\eta}_t)$ in (2.4) is not assumed to be iid. This sequence is only assumed to be a conditionally homoscedastic martingale difference, which allows us to encompass SC models. The analogous of this result was established, in the case of semi-strong univariate GARCH(p, q) models, by Escanciano (2009) as an extension of Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004).

An important class for which Theorem 3.1 applies is the class of DCC models. To our knowledge, no asymptotic estimation results exist in the literature for such models (except the consistency in the corrected "cDCC" version of Aielli (2013)). Stationarity conditions for DCC models have been recently established by Fermanian and Malongo (2014).

3.2. Efficiency loss with respect to the full QMLE?

It can be shown that estimating the volatility coefficients equation by equation does not always entail efficiency loss with respect to the full QML. To see this we compare the efficiency of the full QML estimator (FQMLE) and the EbEE in the bivariate case where, for simplicity, the only unknown coefficients are the parameters of the first volatility. We also assume a constant (and known) correlation matrix. More precisely, consider the model

$$\boldsymbol{\epsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \quad \boldsymbol{H}_t = \begin{pmatrix} \sigma_{1t}^2(\boldsymbol{\theta}_0^{(1)}) & \rho_0\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})\sigma_{2t} \\ \rho_0\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})\sigma_{2t} & \sigma_{2t}^2 \end{pmatrix}$$

where $(\boldsymbol{\eta}_t)$ is as in Model (2.6).

Even if the second equation does not involve the unknown parameter, it conveys information about $\boldsymbol{\theta}_0^{(1)}$ through the correlation. Therefore, it seems that the FQMLE should be more efficient than the EbEE. The next result shows that this is not always the case. The FQMLE of the parameter $\boldsymbol{\theta}_0^{(1)}$ is obtained by minimizing $\sum_{t=1}^n l_t(\boldsymbol{\theta}^{(1)})$ where

$$l_t(\boldsymbol{\theta}^{(1)}) = \log(1-\rho_0^2) + \log\sigma_{1t}^2 + \log\sigma_{2t}^2 + \frac{1}{1-\rho_0^2}\left(\frac{\epsilon_{1t}^2}{\sigma_{1t}^2} + \frac{\epsilon_{2t}^2}{\sigma_{2t}^2} - 2\rho_0\frac{\epsilon_{1t}\epsilon_{2t}}{\sigma_{1t}^2\sigma_{2t}^2}\right),$$

with $\sigma_{1t} = \sigma_{1t}\left(\boldsymbol{\theta}^{(1)}\right)$. Letting $\zeta = \text{Var}\left\{1 - \frac{1}{1-\rho_0^2}\left(\eta_{1t}^* - \rho_0\eta_{2t}^*\right)\eta_{1t}^*\right\}$, it is shown in Appendix B.2 that the FQMLE is asymptotically strictly more efficient than the QMLE based on the first equation if and only if

$$\left(\frac{1-\rho_0^2}{2-\rho_0^2}\right)^2 \zeta < \frac{E\eta_{1t}^{*4}-1}{4}. \tag{3.3}$$

It is interesting to see that the comparison of the two asymptotic variances reduces to a comparison of real numbers. Moreover, these real numbers only depend on the errors distribution, not on the parameters of the volatilities. When $\rho_0 = 0$, the asymptotic variances by the two methods are the same. In the Gaussian case, elementary calculations show that (3.3) holds true: this is not surprising as the FQML coincides with the ML in this case. However, an opposite conclusion may hold for fat tailed distributions.

Roughly speaking, if the errors of a given equation are heavy tailed, it seems preferable to estimate the corresponding volatility without taking the other equations into account.

3.3. *Asymptotic results for strong univariate models*

The asymptotic distribution of the EbEE can be simplified under the assumption that

$$\eta_{kt}^* \text{ is independent from } \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}. \tag{3.4}$$

Moreover, **A8** can be replaced by the weaker assumption

A8*: $E\left|\eta_{kt}^*\right|^4 < \infty$,

and the technical assumptions **A10** on the volatility function can be slightly weakened (see **A10*** in Appendix A). The asymptotic distribution of the EbEE is modified as follows.

THEOREM 3.2. *Under* (3.4) *and the assumptions of Theorem 3.1, with* **A8** *replaced by* **A8*** *and* **A10** *replaced by* **A10****, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left\{0, (E\eta_{kt}^{*4}-1)\boldsymbol{J}_{kk}^{-1}\right\}.$$

It can be noted that Assumption (3.4) is always satisfied in the CCC GARCH case, that is, under (2.6) and $\boldsymbol{R}_t = \boldsymbol{R}$. Interestingly, the next section shows that (3.4) can also be satisfied for certain DCC (GARCH and SC) models.

3.4. Estimating conditional variances in SC models

Because SC models (2.8)-(2.10) satisfy Assumptions (2.4), the volatility parameters $\boldsymbol{\theta}_0^{(k)}$ can be estimated equation by equation, and Theorem 3.1 applies.

We now discuss conditions under which (3.4) holds, in which case the asymptotic covariance matrix of the EbEE simplifies as in Theorem 3.2. The next result shows that when the correlation matrix $\boldsymbol{R}_t^*$ is a function of the latent process $(\boldsymbol{\Delta}_t)$ and when the distribution of $\boldsymbol{\xi}_t$ is spherical, a slightly weaker condition than (3.4) holds. Let $\mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}$ be the σ-field generated by $\{\boldsymbol{\eta}_u^*, u < t\}$.

Proposition 3.1. *Assume that the distribution of $\boldsymbol{\xi}_t$ is spherical and that the sequences $(\boldsymbol{\Delta}_t)$ and $(\boldsymbol{\xi}_t)$ are independent. Then, the SC model (2.8)-(2.10) with $\boldsymbol{R}_t^* = \boldsymbol{R}^*(\Delta_t)$ satisfies*

$$\eta_{kt}^* \text{ is independent from } \mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}. \tag{3.6}$$

Moreover, (η_{kt}^*) is an iid (0,1) sequence.

Remark 3.1. It is worth noting that, under the assumptions of Proposition 3.1, the process $(\boldsymbol{\eta}_t^*)$ is neither independent nor identically distributed in general (even if its components are iid). To see this, consider for example, for $\lambda_1, \lambda_2 \in \mathbb{R}$ and for $k \neq \ell$,

$$\lambda_1 \eta_{kt}^* + \lambda_2 \eta_{\ell t}^* \stackrel{d}{=} \|(\lambda_1 \boldsymbol{e}_k' + \lambda_2 \boldsymbol{e}_\ell')\boldsymbol{R}_t^{*1/2}\|\xi_1 = \{\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2 \boldsymbol{R}_t^*(k,\ell)\}^{1/2}\xi_1,$$

conditionally on $\boldsymbol{R}_t^*$, where $\boldsymbol{e}_k$ denotes the k-th column of $\boldsymbol{I}_m$. The variable in the right-hand side of the latter equality is in general non independent of the past values of $\boldsymbol{\eta}_t^*$, and may also not be stationary (except when $\boldsymbol{R}_t^*(k,\ell)$ is stationary).

Since $\boldsymbol{\eta}_t^* = \boldsymbol{D}_t^{-1}\boldsymbol{\epsilon}_t$ with $\boldsymbol{D}_t \in \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$, it is clear that $\mathcal{F}_{t-1}^{\boldsymbol{\eta}^*} \subset \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$. Therefore (3.4) entails (3.6). Conversely, the equation $\boldsymbol{\epsilon}_t = \boldsymbol{D}_t\boldsymbol{\eta}_t^*$ can be viewed as a GARCH-type model with non iid innovations $(\boldsymbol{\eta}_t^*)$. Under appropriate assumptions on the GARCH recursion defined by $\boldsymbol{D}_t$, the model has a solution of the form $\boldsymbol{\epsilon}_t = \varphi(\boldsymbol{\eta}_t^*, \boldsymbol{\eta}_{t-1}^*, \dots)$ for some measurable function φ. In such a case (3.4) and (3.6) are equivalent, since we have

$$\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}} = \mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}. \tag{3.7}$$

This is illustrated in the following example.

EXAMPLE 3.1 (INFORMATION SETS). Consider the multivariate stationary ARCH(1) model, in which the diagonal elements of $\boldsymbol{H}_t$ have the form

$$\sigma_{it}^2 = \omega_i + \sum_{j=1}^{m} \alpha_{ij}\epsilon_{j,t-1}^2, \qquad \omega_i > 0,\ \alpha_{ij} \geq 0, \quad i,j = 1,\dots,m.$$

Let $\boldsymbol{h}_t = (\sigma_{1t}^2, \dots, \sigma_{mt}^2)'$ and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_m)'$. We have

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \mathbf{A}(\boldsymbol{\eta}_{t-1}^*)\boldsymbol{h}_{t-1},$$

where $\mathbf{A}(\boldsymbol{\eta}_{t-1}^*) = (\alpha_{ij}\eta_{j,t-1}^{*2})_{i,j}$. It follows that

$$\boldsymbol{h}_t = \left(\boldsymbol{I}_m + \sum_{k=1}^{\infty} \mathbf{A}(\boldsymbol{\eta}_{t-1}^*) \dots \mathbf{A}(\boldsymbol{\eta}_{t-k}^*)\right)\boldsymbol{\omega}. \tag{3.8}$$

Under **A1**, the infinite sum is well-defined and is finite componentwise. Otherwise, the norm of $\boldsymbol{h}_t$ would not be finite with probability 1, and this would contradict the strict stationarity of $\boldsymbol{\epsilon}_t$. In view of (3.8), the σ-fields of $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}^*$ coincide, in the sense of (3.7).

A straightforward consequence of Proposition 3.1 and Theorems 3.1-3.2 is the next result.

COROLLARY 3.1. *For Model* (2.8)-(2.10)*, we have strong consistency of* $\hat{\boldsymbol{\theta}}_n^{(k)}$ *under* **A1** *and* **A4**-**A6**. *Under the assumptions of Proposition 3.1 and* (3.7)*, and the additional assumptions* **A7***,* **A8****,* **A9***,* **A10****, the asymptotic normality in* (3.5) *holds.*

4. Estimating conditional and stochastic correlation matrices

Having estimated the individual conditional variances of a vector $(\boldsymbol{\epsilon}_t)$ satisfying (2.4) in a first step, it is generally of interest to estimate the complete conditional variance matrix $\boldsymbol{H}_t$, which reduces to estimating the conditional correlation $\boldsymbol{R}_t$. We first consider the case where $\boldsymbol{R}_t$ is constant, before turning to the estimation of a SC model where the stochastic correlation matrix $\boldsymbol{R}_t^*$ is driven by a Markov chain.

4.1. Estimating generalized CCC models

We consider estimating Model (2.6)-(2.7). The model can be referred to as a Generalized CCC (GCCC) model, as the volatilities are not necessarily specified as functions of the past squared returns (see Section 4.3 for a presentation of classical CCC models).

Let

$$\boldsymbol{\rho} = (R_{21}, \dots, R_{m1}, R_{32}, \dots, R_{m2}, \dots, R_{m,m-1})' = \mathrm{vech}^0(\boldsymbol{R}),$$

denoting by vech^0 the operator which stacks the sub-diagonal elements (excluding the diagonal) of a matrix. The global parameter, denoted

$$\boldsymbol{\vartheta} = (\boldsymbol{\theta}^{(1)'}, \dots, \boldsymbol{\theta}^{(m)'}, \boldsymbol{\rho}')' := (\boldsymbol{\theta}', \boldsymbol{\rho}')' \in \mathbb{R}^d \times [-1,1]^{m(m-1)/2}, \quad d = \sum_{k=1}^m d_k,$$

belongs to the compact parameter set $\boldsymbol{\Theta} = \prod_{k=1}^m \boldsymbol{\Theta}_k \times [-1,1])^{m(m-1)/2}$. The true parameter value is

$$\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0^{(1)'}, \dots, \boldsymbol{\theta}_0^{(m)'}, \boldsymbol{\rho}_0')' := (\boldsymbol{\theta}_0', \boldsymbol{\rho}_0')'.$$

We now consider a two-step method for estimating $\boldsymbol{\vartheta}_0$ which can be summarized as follows:

(a) Estimation of $\boldsymbol{\theta}_0^{(k)}$, equation-by-equation, in the individual GARCH-type models (2.5) and extraction of the residuals of the k-th equation, $\hat{\eta}^*_{kt} = \tilde{\sigma}_{kt}^{-1}(\hat{\boldsymbol{\theta}}^{(k)})\epsilon_{kt}$;

(b) Computation of the empirical correlation matrix

$$\hat{\boldsymbol{R}}_n = \frac{1}{n}\sum_{t=1}^n \hat{\boldsymbol{\eta}}_t^* \left(\hat{\boldsymbol{\eta}}_t^*\right)',$$

where $\hat{\boldsymbol{\eta}}_t^*$ is the vector of residuals of the m equations.

Let

$$\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\boldsymbol{\theta}}_n' := (\hat{\boldsymbol{\theta}}_n^{(1)'}, \dots, \hat{\boldsymbol{\theta}}_n^{(m)'}), \hat{\boldsymbol{\rho}}_n'\right)', \qquad \hat{\boldsymbol{\rho}}_n = \mathrm{vech}^0(\hat{\boldsymbol{R}}_n).$$

THEOREM 4.1. *For the GCCC model* (2.6)-(2.7), *if* **A1**-**A6** *hold, then*

$$\hat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0, \quad a.s. \quad as\ n \to \infty.$$

For the asymptotic normality, we introduce the following notations. Let the $d \times d$ matrix $\boldsymbol{J}^* = ((\kappa^*_{k\ell} - 1)\boldsymbol{J}_{k\ell})$ where $\kappa^*_{k\ell} = E\left(\eta_{kt}^{*2}\eta_{\ell t}^{*2}\right)$, for $k, \ell = 1, \dots, m$, and $\boldsymbol{J}_{k\ell} = E\left(\boldsymbol{d}_{kt}\boldsymbol{d}_{\ell t}'\right)$. Let, for $\boldsymbol{J}_0 = \mathrm{diag}(\boldsymbol{J}_{11}, \dots, \boldsymbol{J}_{mm})$ in bloc-matrix notation,

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{J}_0^{-1}\boldsymbol{J}^*\boldsymbol{J}_0^{-1} = \left((\kappa^*_{k\ell} - 1)\boldsymbol{J}_{kk}^{-1}\boldsymbol{J}_{k\ell}\boldsymbol{J}_{\ell\ell}^{-1}\right).$$

Let also $\boldsymbol{d}_t = (\boldsymbol{d}_{1t}', \dots, \boldsymbol{d}_{mt}')' \in \mathbb{R}^d$, $\boldsymbol{\Omega}_k = E\boldsymbol{d}_{kt}$ and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1', \dots, \boldsymbol{\Omega}_m')' \in \mathbb{R}^d$. Let $\boldsymbol{\Gamma} = \mathrm{var}\left(\mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\}\right)$. For $x \in \mathbb{R}^m$, let the $d \times d$ matrices $F(x) = \mathrm{diag}\{(1 - x_1^2)\boldsymbol{j}_1, \dots, (1 -$

$x_m^2)\boldsymbol{j}_m\}$, where $j_k = (1, \dots, 1) \in \mathbb{R}^{d_k}$, and $\boldsymbol{A}_{k\ell} = E\{\eta_{kt}^* \eta_{\ell t}^* F(\boldsymbol{\eta}_t^*)\}$. Let, for $k, \ell = 2, \dots, m$, the $d \times d$ matrix $\mathbf{M}_{k,\ell-1} = \text{diag}\left(\mathbf{M}_{k,\ell-1}^{(1)}, \dots, \mathbf{M}_{k,\ell-1}^{(m)}\right)$ where

$$\mathbf{M}_{k,\ell-1}^{(i)} = \begin{cases} \mathbf{0}_{d_i \times d_i} & \text{if } i \neq k \quad \text{and} \quad i \neq \ell \\ R_{k,\ell-1} \boldsymbol{I}_{d_i} & \text{otherwise.} \end{cases}$$

Let the $d \times dm(m-1)/2$ matrices $\boldsymbol{A} = (\boldsymbol{A}_{21} \dots \boldsymbol{A}_{m1} \;\; \boldsymbol{A}_{32} \dots \boldsymbol{A}_{m,m-1})$ and $\boldsymbol{M} = (\boldsymbol{M}_{21} \dots \boldsymbol{M}_{m1} \; \boldsymbol{M}_{32} \dots \boldsymbol{M}_{m,m-1})$. Let the $d \times m(m-1)/2$ matrices

$$\boldsymbol{L} = \boldsymbol{A}(\boldsymbol{I}_{m(m-1)/2} \otimes \boldsymbol{\Omega}), \qquad \boldsymbol{\Lambda} = \boldsymbol{M}(\boldsymbol{I}_{m(m-1)/2} \otimes \boldsymbol{\Omega}).$$

Let

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\rho}} = -\frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{\Lambda} - \boldsymbol{J}_0^{-1}\boldsymbol{L}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\rho}} = \frac{1}{4}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{\Lambda} + \frac{1}{2}\left(\boldsymbol{\Lambda}'\boldsymbol{J}_0^{-1}\boldsymbol{L} + \boldsymbol{L}'\boldsymbol{J}_0^{-1}\boldsymbol{\Lambda}\right) + \boldsymbol{\Gamma}.$$

We need an additional assumption.

A13: The m components of $\boldsymbol{\eta}_t$ are mutually independent random variables.

THEOREM 4.2. *For the GCCC model* (2.6)-(2.7)*, if* **A1**-**A13** *hold, for* $k = 1, \dots, m$, *and* $\boldsymbol{\rho}_0 \in (-1,1)^{m(m-1)/2}$*, then*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \\ \sqrt{n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \boldsymbol{\Sigma} := \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} & \boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\rho}} \\ \boldsymbol{\Sigma}'_{\boldsymbol{\theta}\boldsymbol{\rho}} & \boldsymbol{\Sigma}_{\boldsymbol{\rho}} \end{pmatrix}\right\},$$

and $\boldsymbol{\Sigma}$ *is a non-singular matrix.*

REMARK 4.1. Even though the components of $\boldsymbol{\theta}_0$ are estimated equation by equation, the components of $\hat{\boldsymbol{\theta}}_n$ are not asymptotically independent in general. More precisely, it can be seen that

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} \text{ is diagonal if } \text{Cov}(\eta_{kt}^{*2}, \eta_{\ell t}^{*2}) = 0 \;\text{ for any } k \neq \ell.$$

REMARK 4.2. In the asymptotic variance $\boldsymbol{\Sigma}_{\boldsymbol{\rho}}$ of $\hat{\boldsymbol{\rho}}_n$, the first two matrices in the sum reflect the effect of the estimation of $\boldsymbol{\theta}_0$, while the remaining matrix, $\boldsymbol{\Gamma}$, is independent of $\boldsymbol{\theta}_0$. A limit case is when the components of $\boldsymbol{\eta}_t^*$ are serially independent, that is when $\boldsymbol{\eta}_t^* = \boldsymbol{\eta}_t$ and $\boldsymbol{R}$ is the identity matrix. Then, straightforward computation shows that $\boldsymbol{L} = \boldsymbol{\Lambda} = \mathbf{0}$ and thus

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{m(m-1)/2} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \text{diag}((\kappa_{11}^* - 1)\boldsymbol{J}_{11}^{-1}, \dots, (\kappa_{mm}^* - 1)\boldsymbol{J}_{mm}^{-1})$$

in bloc-matrix notation.

Remark 4.3. It can be seen from the proof that Assumption **A13** is only used to show that $\boldsymbol{\Sigma}$ is non singular.

Remark 4.4. It is worthnoting that all the matrices involved in the asymptotic covariance matrix $\boldsymbol{\Sigma}$ take the form of expectations. A simple estimator of $\boldsymbol{\Sigma}$ is thus obtained by replacing those expectations by their sample counterparts. For instance, it can be shown that a consistent estimator of $\boldsymbol{A}_{k\ell}$ is

$$\hat{\boldsymbol{A}}_{k\ell} = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_{kt}^{*}\hat{\eta}_{\ell t}^{*}F(\hat{\boldsymbol{\eta}}_t^{*}).$$

4.2. Estimating stochastic correlations driven by an hidden Markov chain

A natural extension of the generalized CCC model is obtained by allowing the matrix $\boldsymbol{R}_t^*$ to be driven by a Markov chain. This extension was studied by Pelletier (2006). Assume that $(\boldsymbol{\epsilon}_t)$ is generated by Model (2.8) with

$$\boldsymbol{R}_t^* = \boldsymbol{R}^*(\Delta_t), \text{ where } (\Delta_t) \text{ is a Markov chain on } \mathcal{E} = \{1,\dots,N\}. \tag{4.2}$$

Note that the Markov chain is not observed but the number of states, N, is assumed to be known. Denoting by $p(i,j) = P(\Delta_t = j \mid \Delta_{t-1} = i)$ the transition probabilities of the Markov chain, the parameter vector is now denoted

$$\begin{aligned}\boldsymbol{\zeta} &= (\boldsymbol{\theta}^{(1)'},\dots,\boldsymbol{\theta}^{(m)'},\boldsymbol{\rho}'(1),\dots,\boldsymbol{\rho}'(N),\boldsymbol{p}')' \\ &:= (\boldsymbol{\theta}',\boldsymbol{\rho}',\boldsymbol{p}')' \in \mathbb{R}^d \times [-1,1]^{Nm(m-1)/2} \times [0,1]^{N(N-1)},\end{aligned}$$

where $\boldsymbol{p} = (p(1,2),p(1,3),\dots,p(1,N),p(2,2),\dots,p(N,N))'$ and $\boldsymbol{\rho}(i) = \mathrm{vech}^0\{\boldsymbol{R}(i)\}$ for $i = 1,\dots,N$.

A common approach to estimating Hidden Markov Models (HMM) is maximum likelihood estimation (MLE). There is a vast literature on the estimation of HMM. To mention just a few, see for instance Baum (1972), Baum and Petrie (1966), Francq and Roussignol (1995), Francq, Roussignol and Zakoïan (2001) and the overviews by Cappé, Ryden and Moulines (2005), and Frühwirth-Schnatter (2005). In this paper, we do not use the full maximum likelihood method which is generally intractable, in particular when the regimes are not Markovian (that is, when the conditional variances σ_{kt}^2 do not depend on a finite number of past values of $\boldsymbol{\epsilon}_t$). Instead, we follow a two-step approach: having estimated $\boldsymbol{\theta}_0$ in the first step, we apply the maximum likelihood on the standardized residuals to estimate the remaining parameters in a second step. To this aim, we need to specify the errors distribution. We assume that

A14: the sequences (Δ_t) and (ξ_t) are mutually independent.

A15: the Markov chain (Δ_t) is stationary, irreducible and aperiodic.

A16: $\boldsymbol{\xi}_t$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_m$.

We start by considering the case where the GARCH part is absent in the dynamics (*i.e.* $\boldsymbol{D}_t = \boldsymbol{I}_m$ in (2.8)). Let $\boldsymbol{\eta}_1^*, \dots, \boldsymbol{\eta}_n^*$ be observations of the HMM model

$$\boldsymbol{\eta}_t^* = \{\boldsymbol{R}^*(\Delta_t)\}^{1/2}\boldsymbol{\xi}_t, \tag{4.3}$$

with unknown parameter $\boldsymbol{\vartheta}_0 = (\boldsymbol{\rho}_0', \boldsymbol{p}_0')'$. The likelihood of the model is obtained by summing, over all possible paths of the Markov chain, the probability densities at the points $(\boldsymbol{\eta}_1^*, \dots, \boldsymbol{\eta}_n^*)$:

$$L_n(\boldsymbol{\vartheta}) = \sum_{\{e_1,\dots,e_n\}\in\mathcal{E}^n} \pi(e_1)\left\{\prod_{t=2}^n p(e_{t-1}, e_t)\right\}\left\{\prod_{t=1}^n f_{\boldsymbol{\eta}_t^*}(e_t)\right\}$$

where $\pi(1), \dots, \pi(N)$ denote the stationary probabilities of the chain and, denoting by $|\boldsymbol{A}|$ the determinant of a square matrix $\boldsymbol{A}$, for $\boldsymbol{x} \in \mathbb{R}^m$,

$$f_{\boldsymbol{x}}(i) = \frac{1}{(2\pi)^{m/2}}|\boldsymbol{R}^*(i)|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{R}^*(i)^{-1}\boldsymbol{x}\right\}.$$

Direct computation of the likelihood based on this formula rapidly becomes intractable as the sample size, n, increases. However, the likelihood can be expressed as a sum of products of matrices, using the following notations. For any function $f:\ \mathcal{E} \to \mathbb{R}$, let the matrix

$$\mathbb{P}(f) = \begin{pmatrix} p(1,1)f(1) & \cdots & p(N,1)f(1) \\ \vdots & & \vdots \\ p(1,d)f(N) & \cdots & p(N,N)f(N) \end{pmatrix}, \quad \text{and the vector } \Pi(f) = \begin{pmatrix} \pi(1)f(1) \\ \vdots \\ \pi(N)f(N) \end{pmatrix}.$$

Then, following Francq and Roussignol (1997), the likelihood can be written as

$$L_n(\boldsymbol{\vartheta}) = \boldsymbol{e}'\prod_{t=2}^n \mathbb{P}(f_{\boldsymbol{\eta}_t^*})\Pi(f_{\boldsymbol{\eta}_1^*}), \tag{4.4}$$

where $\boldsymbol{e} = (1, \dots, 1)' \in \mathbb{R}^N$.

The parameter space Θ^* for $\boldsymbol{\vartheta}$ is defined as a compact subset $\Theta^*_{\boldsymbol{\rho}} \times \Theta^*_{\boldsymbol{p}}$ of $[-1,1]^{Nm(m-1)/2} \times [0,1]^{N(N-1)}$ which contains the true value $\boldsymbol{\vartheta}_0$ and is compatible with Assumption **A15** (that is, the Markov chain is stationary, irreducible and aperiodic for any parameter value $\boldsymbol{p}$). It is also necessary to constrain the parameter space so that the parameter be identifiable. Yakowitz and Spragins (1968) showed that finite mixtures of m-dimensional Gaussian distributions (with distinct pairs $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ of mean and covariance matrix) are identifiable. In Model (4.3), the multivariate Gaussian distributions corresponding to the different regimes of the Markov chain are centered. A way to ensure that the variances be different and cannot be permuted is to use the lexicographical order. Therefore, we assume that for any $\boldsymbol{\rho} \in \Theta^*_{\boldsymbol{\rho}}$,

$$\boldsymbol{\rho}(1) \prec \boldsymbol{\rho}(2) \prec \cdots \prec \boldsymbol{\rho}(N),$$

in the sense of the lexicographical order[2].

Let $(\hat{\boldsymbol{\vartheta}}_n)$ be a sequence such that

$$L_n(\hat{\boldsymbol{\vartheta}}_n) = \sup_{\boldsymbol{\vartheta} \in \Theta^*} L_n(\boldsymbol{\vartheta}). \tag{4.5}$$

THEOREM 4.3. *For the Hidden Markov DCC model* (4.3)*, if* **A14, A15, A16** *hold, then*

$$\hat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0, \quad a.s. \quad as\ n \to \infty.$$

It is possible to obtain the MLE $\hat{\boldsymbol{\vartheta}}_n$ from (4.5), by numerical optimization of the likelihood computed from (4.4). It is however numerically more efficient to use the filter proposed by Hamilton (1989) for computing and optimizing the log-likelihood of an HMM model. The log-likelihood of the model (4.3) is given by

$$\log L_n(\boldsymbol{\vartheta}) = \sum_{t=1}^{n} \log \mathbf{1}' \left\{ \boldsymbol{\pi}_{t|t-1} \odot \boldsymbol{\phi}(\boldsymbol{\eta}_t^*) \right\},$$

where all the elements of the vector $\mathbf{1}$ are equal to 1, $\odot$ denotes Hadamard's product of matrices, $\boldsymbol{\phi}(\boldsymbol{\eta}_t^*) = \left(f_{\boldsymbol{\eta}_t^*}(1), \ldots, f_{\boldsymbol{\eta}_t^*}(N)\right)'$, and

$$\boldsymbol{\pi}_{t|s} \quad = \quad \left(P(\Delta_t = 1 | \boldsymbol{\eta}_s^*, \ldots, \boldsymbol{\eta}_1^*), \ldots, P(\Delta_t = d | \boldsymbol{\eta}_s^*, \ldots, \boldsymbol{\eta}_1^*)\right)'.$$

Let $\boldsymbol{P}$ be the matrix of the transition probabilities, with $p(i,j)$ as element of the i-th row and j-th column. Let also $\boldsymbol{\pi}_0 = (\pi(1), \ldots, \pi(N))'$. Adapting Hamilton's EM algorithm

[2]For two vectors $\boldsymbol{x} = (x_i)$ and $\boldsymbol{y} = (y_i)$ of the same dimension, we have $\boldsymbol{x} \prec \boldsymbol{y}$ if and only if there exists $i > 0$ such that for all $j < i$ we have $x_j = y_j$ and $x_i < y_i$.

to our framework, the maximum likelihood can be obtained by starting with initial values for $\boldsymbol{\pi}_0$ and $\boldsymbol{\vartheta}$, and iterating until convergence the following steps:

(a) Set $\boldsymbol{\pi}_{1|0} = \boldsymbol{\pi}_0$ and

$$\boldsymbol{\pi}_{t|t} = \frac{\boldsymbol{\pi}_{t|t-1} \odot \boldsymbol{\phi}(\boldsymbol{\eta}_t^*)}{\mathbf{1}' \left\{ \boldsymbol{\pi}_{t|t-1} \odot \boldsymbol{\phi}(\boldsymbol{\eta}_t^*) \right\}}, \quad \boldsymbol{\pi}_{t+1|t} = \boldsymbol{P}' \boldsymbol{\pi}_{t|t}, \quad \text{for } t = 1, \dots, n.$$

(b) Compute the smoothed probabilities $\boldsymbol{\pi}_{t|n}(i) = P(\Delta_t = i \mid \boldsymbol{\eta}_1^*, \dots, \boldsymbol{\eta}_n^*)$ by

$$\boldsymbol{\pi}_{t-1|n}(i) \quad = \quad \sum_{j=1}^{d} \frac{p(i,j) \boldsymbol{\pi}_{t-1|t-1}(i) \boldsymbol{\pi}_{t|n}(j)}{\boldsymbol{\pi}_{t|t-1}(j)} \quad \text{for } t = n, n-1, \dots, 2,$$

and $\boldsymbol{\pi}_{t-1,t|n}(i,j) = P(\Delta_{t-1} = i, \Delta_t = j \mid \boldsymbol{\eta}_1^*, \dots, \boldsymbol{\eta}_n^*)$ by

$$\boldsymbol{\pi}_{t-1,t|n}(i,j) \quad = \quad \frac{p(i,j) \boldsymbol{\pi}_{t-1|t-1}(i) \boldsymbol{\pi}_{t|n}(j)}{\boldsymbol{\pi}_{t|t-1}(j)}.$$

(c) Replace the previous values of the parameters by $\boldsymbol{\pi}_0 = \boldsymbol{\pi}_{1|n}$,

$$p(i,j) = \frac{\sum_{t=2}^{n} \boldsymbol{\pi}_{t-1,t|n}(i,j)}{\sum_{t=2}^{n} \boldsymbol{\pi}_{t-1|n}(i)}$$

and, denoting by $\mathcal{R}$ the space of the $m \times m$ symmetric positive definite matrices, compute

$$\boldsymbol{R}^*(i) = \arg\min_{\boldsymbol{R} \in \mathcal{R}} \log |\boldsymbol{R}| + \operatorname{Tr} \left\{ \boldsymbol{R}^{-1} \boldsymbol{\Sigma}(i) \right\} \tag{4.6}$$

where

$$\boldsymbol{\Sigma}(i) = \frac{1}{\sum_{t=1}^{n} \boldsymbol{\pi}_{t|n}(i)} \sum_{t=1}^{n} \boldsymbol{\eta}_t^* (\boldsymbol{\eta}_t^*)' \boldsymbol{\pi}_{t|n}(i). \tag{4.7}$$

In the standard version of Hamilton's EM algorithm, the unknown coefficients are the variance matrices $\boldsymbol{\Sigma}(i)$ of the Gaussian distributions, and the M step consists in maximizing

$$\sum_{i=1}^{N} \sum_{t=1}^{n} \log f_{\boldsymbol{\eta}_t^*}(i) \boldsymbol{\pi}_{t|n}(i) = \frac{-1}{2} \sum_{i=1}^{N} \sum_{t=1}^{n} \left\{ \log |\boldsymbol{\Sigma}(i)| + (\boldsymbol{\eta}_t^*)' \boldsymbol{\Sigma}(i)^{-1} \boldsymbol{\eta}_t^* \right\} \boldsymbol{\pi}_{t|n}(i)$$

with respect to the $\boldsymbol{\Sigma}(i)$'s. The solution of this optimization problem is given explicitly by (4.7). The matrix $\boldsymbol{R}^*(i)$ defined in (4.6) can thus be interpreted as the correlation matrix which is the closest to the covariance matrix provided by the EM algorithm.

In practice, when a GARCH part is present (*i.e.* $\boldsymbol{D}_t \neq \boldsymbol{I}_m$ in (2.8)), the innovations $\boldsymbol{\eta}_t^*$'s are not available. Note however that, under the assumptions of Theorem 3.1, the equation-by-equation GARCH estimator $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. The EM algorithm can then be applied to the residuals $\hat{\boldsymbol{\eta}}_t^* = \tilde{\boldsymbol{\eta}}_t^*(\hat{\boldsymbol{\theta}}_n) = \tilde{\boldsymbol{D}}_t^{-1}(\hat{\boldsymbol{\theta}}_n) \boldsymbol{\epsilon}_t, \quad t = 1, \dots n.$

4.3. Time complexity comparison of the EbEE and the full QMLE

Bollerslev (1990) introduced the CCC-GARCH(p,q) model

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \sum_{i=1}^{q} \boldsymbol{A}_i \underline{\boldsymbol{\epsilon}}_{t-i} + \sum_{j=1}^{p} \boldsymbol{B}_j \boldsymbol{h}_{t-j}$$

where $\boldsymbol{h}_t = \left(\sigma_{1t}^2, \cdots, \sigma_{mt}^2\right)'$, $\underline{\boldsymbol{\epsilon}}_t = \left(\epsilon_{1t}^2, \cdots, \epsilon_{mt}^2\right)'$, $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are diagonal $m \times m$ matrices with positive coefficients and $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_m)'$ is a vector of strictly positive coefficients. An extended version of this model, called the *Extended* CCC model by He and Teräsvirta (2004), relaxes the assumption that the matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are diagonal. Let us compare the computation time of the EbEE with that of the FQMLE in the case of an extended CCC-GARCH$(1,1)$ model of dimension m, in which $\boldsymbol{A}_1 = (\alpha_{ij})$ and $\boldsymbol{B}_1 = \mathrm{diag}(\beta_1, \ldots, \beta_m)$. The conditional variance of the k-th component of this model is thus equal to

$$\sigma_{kt}^2 = \omega_k + \sum_{j=1}^{m} \alpha_{kj} \epsilon_{j,t-1}^2 + \beta_k \sigma_{k,t-1}^2.$$

The EbEE of all the parameters of the model requires m estimations of univariate GARCH-type models with $m+2$ parameters, plus the computation of the empirical correlation of the EbE residuals. The full QMLE requires the optimization of a function of the $m^2 + 2m + m(m-1)/2$ parameters of the model. Because the time complexity of an optimization generally grows rapidly with the dimension of the objective function, the full QMLE should be much more costly than the EbEE in terms of computation time. Table 1 compares the effective computation times required by the two estimators as a function of the dimension m, for the exchange rate series that will be studied in Section 6 below. These time series have length $n = 2081$. As expected, the comparison is clearly in favor of the EbEE. Note that these computation times have been obtained using a single processor. Since the EbEE is clearly easily parallelizable (using one processor for each of the m optimizations), the advantage of the EbEE should be even more pronounced with a multiprocessing implementation.

5. Testing for adequacy of particular MGARCH models

Theorem 3.1 can be used for testing the adequacy of a particular class of MGARCH models, preliminary to its estimation. Indeed, most commonly used MGARCH specifications imply

Table 1. Computation time of the two estimators (CPU time in seconds)

	dimension m				
	2	3	4	5	6
Estimator					
EbEE	15.59	28.50	43.91	70.90	98.39
FQMLE	101.41	443.34	870.04	1182.22	1515.58

strong restrictions on the volatility of the individual components. Let us focus on the class of BEKK models.

For simplicity, consider the simplest model of this form, namely the bivariate BEKK-GARCH(1,1) model given by

$$\boldsymbol{\epsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \qquad \boldsymbol{H}_t = \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{H}_{t-1}, \tag{5.1}$$

where $(\boldsymbol{\eta}_t)$ is an iid $\mathbb{R}^2$-valued centered sequence with $E\boldsymbol{\eta}_t\boldsymbol{\eta}'_t = \boldsymbol{I}_2$, $\boldsymbol{A} = (a_{ij})_{1\leq i,j\leq 2}$ and $\boldsymbol{B} = \mathrm{diag}(b_1, b_2)$ with $b_1, b_2 \geq 0$, and $\boldsymbol{\Omega}$ is a positive definite 2×2 matrix. It follows that the diagonal entries of $\boldsymbol{H}_t$ are given by

$$\begin{cases} h_{11,t} = \omega_{11} + a_{11}^2\epsilon_{1,t-1}^2 + 2a_{11}a_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{12}^2\epsilon_{2,t-1}^2 + b_1 h_{11,t-1}, \\ h_{22,t} = \omega_{22} + a_{21}^2\epsilon_{1,t-1}^2 + 2a_{21}a_{22}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{22}^2\epsilon_{2,t-1}^2 + b_2 h_{22,t-1}. \end{cases}$$

Letting $\boldsymbol{\theta}_0^{(k)} = (\omega_{kk}, a_{k1}^2, 2a_{k1}a_{k2}, a_{k2}^2)'$ for $k = 1, 2$, the validity of this model can be studied by estimating Model (2.5) for each component of $\boldsymbol{\epsilon}_t$, with

$$\sigma_{kt}^2 = \theta_{01}^{(k)} + \theta_{02}^{(k)}\epsilon_{1,t-1}^2 + \theta_{03}^{(k)}\epsilon_{1,t-1}\epsilon_{2,t-1} + \theta_{04}^{(k)}\epsilon_{2,t-1}^2 + \theta_{05}^{(k)}\sigma_{k,t-1}^2, \quad k = 1, 2, \tag{5.2}$$

under the positivity constraints $\theta_{01}^{(k)} > 0, \quad \theta_{0i}^{(k)} \geq 0, i = 2, 5$. The restrictions implied by the BEEK-GARCH(1,1) model (5.1) are of the form:

$$H_0^{(k)}: \qquad \theta_{03}^{(k)} = 2\sqrt{\theta_{02}^{(k)}\theta_{04}^{(k)}}, \quad k = 1, 2.$$

Let

$$\boldsymbol{\Theta}^{(k)} = \boldsymbol{\Theta}_k^* \cap \left\{\boldsymbol{\theta}^{(k)}; \theta_3^{(k)} \in \left[0, 2\sqrt{\theta_2^{(k)}\theta_4^{(k)}}\right]\right\},$$

where $\boldsymbol{\Theta}_k^*$ is a compact subset of $\{\theta_1^{(k)} > 0, \theta_i^{(k)} \geq 0$, for $i = 2, 3, 4$ and $\theta_5^{(k)} \in [0, 1)\}$. Note that, under $H_0^{(k)}$, the true parameter value is at the boundary of the parameter set.

THEOREM 5.1. *Let the spectral radius of $\boldsymbol{A}+\boldsymbol{B}$ be less than 1, and let $a_{11}a_{12}>0$, $a_{21}a_{22}>0$. Let $\boldsymbol{\eta}_1$ admit, with respect to the Lebesgue measure on $\mathbb{R}^2$, a positive density around 0, and suppose that $E\,|\eta_{kt}|^{4(1+\delta)}<\infty$, for $k=1,2$ and some $\delta>0$. Let $\boldsymbol{\theta}_0^{(k)}$ belong to the interior of $\boldsymbol{\Theta}_k^*$ for $k=1,2$.*

Let $(\boldsymbol{\epsilon}_t)$ be the strictly stationary solution of Model (5.1). Let the Wald statistic for the hypothesis $H_0^{(k)}$,

$$\boldsymbol{W}_n^{(k)}=\frac{n\left\{\hat{\theta}_{n3}^{(k)}-2\sqrt{\hat{\theta}_{n2}^{(k)}\hat{\theta}_{n4}^{(k)}}\right\}^2}{\boldsymbol{X}_n'\hat{\boldsymbol{J}}_{kk}^{-1}\hat{\boldsymbol{I}}_{kk}\hat{\boldsymbol{J}}_{kk}^{-1}\boldsymbol{X}_n},\qquad \text{where}\quad \hat{\boldsymbol{\theta}}_n^{(k)}=(\hat{\theta}_{n1}^{(k)},\dots,\hat{\theta}_{n5}^{(k)})',$$

$$\boldsymbol{X}_n=\left(0,\sqrt{\hat{\theta}_{n4}^{(k)}/\hat{\theta}_{n2}^{(k)}},-1,\sqrt{\hat{\theta}_{n2}^{(k)}/\hat{\theta}_{n4}^{(k)}},0\right)',\ \hat{\eta}_{kt}^*=\epsilon_{kt}/\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})\ \text{and}$$

$$\hat{\boldsymbol{J}}_{kk}=\frac{1}{n}\sum_{t=1}^n\hat{\boldsymbol{d}}_{kt}\hat{\boldsymbol{d}}_{kt}',\quad \hat{\boldsymbol{I}}_{kk}=\frac{1}{n}\sum_{t=1}^n\{\hat{\eta}_{kt}^{*4}-1\}\hat{\boldsymbol{d}}_{kt}\hat{\boldsymbol{d}}_{kt}',\quad \hat{\boldsymbol{d}}_{kt}=\frac{1}{\tilde{\sigma}_{kt}^2(\hat{\boldsymbol{\theta}}_n)}\frac{\partial\tilde{\sigma}_{kt}^2(\hat{\boldsymbol{\theta}}_n^{(k)})}{\partial\boldsymbol{\theta}^{(k)}}.$$

Then, $\boldsymbol{W}_n^{(k)}$ asymptotically follows a mixture of the χ^2 distribution with one degree of freedom and the Dirac measure at 0:

$$\boldsymbol{W}_n^{(k)}\xrightarrow{\mathcal{L}}\frac{1}{2}\chi^2(1)+\frac{1}{2}\delta_0\quad \text{as } n\to\infty.$$

In view of this result, testing $H_0^{(k)}$ at the asymptotic level $\alpha\in(0,1/2)$ can thus be achieved by using the critical region $\{\boldsymbol{W}_n^{(k)}>\chi^2_{1-2\alpha}(1)\}$.

6. Illustrations

We present two applications. The first one shows that the two-step EbEE can easily estimate a CCC-GARCH model, even if the different components of the multivariate series of returns are not observed simultaneously. In that case, the individual volatilities have however to follow pure GARCH models. In the second application, the individual volatilities are augmented GARCH models and the conditional correlation displays several regimes. The second application also illustrates the specification test based on the EbEE.

6.1. An application to world stock market indices

From the Yahoo Finance Website `http://finance.yahoo.com/`, we downloaded the whole set of the major World indices. We kept for these series the names given by Yahoo. We took the daily data available over the period from 1990-01-01 to 2013-04-22, and we eliminated a few series with too few observations. We then obtained a total number of 25 series: 5 for Americas, 11 for Asia-Pacific, 8 for Europe and 1 for Middle East. Because some series do not cover the entire period and the working days are not the same for all the financial markets, the number n of observations varies a lot, from $n = 2157$ for the series "NZ50" to $n = 6040$ for "AEX.AS". We corrected the "MERV" series for the stock spilt that occurred in Brazil on 1997-03-11, and we started at 1990-08-02 for the series "GD.AT" because of the presence of unexpected variations before this date. On each of the 25 series, we fitted PGARCH(1,1) models of the form

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \alpha_+ (\epsilon_{t-1}^+)^\delta + \alpha_- (-\epsilon_{t-1}^-)^\delta + \beta \sigma_{t-1}^\delta \end{cases} \tag{6.1}$$

where $x^+ = \max(x, 0)$, $x^- = \min(x, 0)$, $\alpha_+ \geq 0$, $\alpha_- \geq 0$, $\beta \in [0, 1)$, $\omega > 0$, and $\delta > 0$. As shown by Hamadeh and Zakoian (2011), the effective estimation of the parameter δ is an issue. The quasi-likelihood in the direction of δ being often relatively flat, the QML estimation of this parameter is imprecise and considerably slows down the optimization procedure. For this reason we decided to perform the QML optimization on only 4 values of this parameter: $\delta \in \{0.5, 1, 1.5, 2\}$. For each of the 4 values of δ, the remainder parameter $\boldsymbol{\theta} = (\omega, \alpha_+, \alpha_-, \beta)'$ is estimated by QML. Following the (quasi-)likelihood principle, the selected values of δ and the final estimated value of $\boldsymbol{\theta}$ maximize the QML over the 4 optimizations.

Table 2 displays the estimated PGARCH(1,1) models for each series, the estimated standard deviation into parentheses, and the selected value of δ in the last column. For all series, one can see a strong leverage effect ($\alpha_- > \alpha_+$) which means that negative returns tend to have an higher impact on the future volatility than positive returns of the same magnitude.

Table 3 gives an empirical estimate $\hat{\boldsymbol{R}}$ of the correlation matrix $\boldsymbol{R}$ of the residuals of the 25 PGARCH(1,1) equations. Because there are numerous missing values, due to the fact that the series are not always observed at the same dates, we used the `R` function `cor()` with the option "use=pairwise.complete.obs", which means that the correlation between each pair of variables is computed using all complete pairs of observations on those variables.

A principal component analysis (PCA) has been performed on the matrix $\hat{\boldsymbol{R}}$. The percentage of variance explained by the first four principal components are respectively

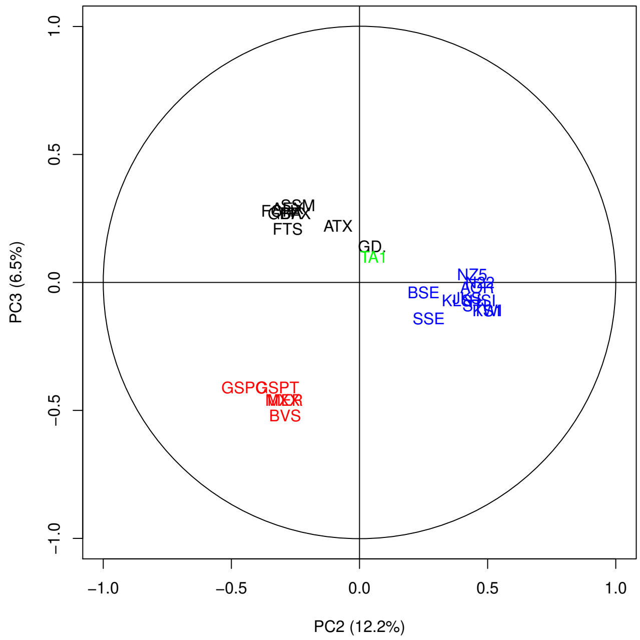

Figure 1. Factorial plan PC2-PC3.

34.6%, 12.2%, 6.5% and 3.8%. Table 4 gives the so-called loading matrix, that is the correlation between the variables and the factors. From this table, it is clear that the first principal component PC1 is a scaling factor. PC1 is negatively correlated with all the series of returns. Noting that, in (6.1), the signs of ϵ_t and η_t are the same, the PC1 factor thus opposes the days where the markets are globally profitable to days where the markets go down. Therefore, we can interpret PC1 as the global trend of the World markets (with the negative sign for PC1 when the returns are globally positive). The second factor PC2 opposes the American and European to the Asian markets, whereas PC3 opposes the European and American markets (see Figure 1 for a graphical illustration). These relationships are certainly related to the opening hours of the different markets.

6.2. *An application to exchange rates*

We now consider returns series of the daily exchange rates of the Canadian Dollar (CAD), the Swiss Franc (CHF), the Chinese Yuan (CNY), the British Pound (GBP), the Japanese Yen (JPY) and the American Dollar (USD) with respect to the Euro. The observations have been downloaded from the website `http://www.ecb.int/home/html/index.en.html`,

Table 2. PGARCH(1,1) models fitted by EbEE on daily returns of the major World stock indices. The estimated standard deviation are displayed into parentheses. The last column gives the selected value of the power δ.

	$\widehat{\omega}$	$\widehat{\alpha}_+$	$\widehat{\alpha}_-$	$\widehat{\beta}$	$\widehat{\delta}$
MERV	0.151 (0.002)	0.063 (0.002)	0.151 (0.001)	0.858 (0.004)	2
BVSP	0.077 (0.001)	0.068 (0.001)	0.138 (0.002)	0.884 (0.002)	2
GSPTSE	0.012 (0.009)	0.046 (0.002)	0.109 (0.004)	0.926 (0.007)	1
MXX	0.032 (0.003)	0.044 (0.001)	0.167 (0.002)	0.896 (0.004)	1.5
GSPC	0.016 (0.006)	0.000 (0.002)	0.134 (0.003)	0.927 (0.004)	1.5
AORD	0.023 (0.007)	0.030 (0.002)	0.131 (0.003)	0.910 (0.006)	1
SSEC	0.031 (0.010)	0.082 (0.004)	0.123 (0.003)	0.904 (0.012)	1
HSI	0.029 (0.008)	0.049 (0.003)	0.120 (0.003)	0.916 (0.009)	1
BSESN	0.055 (0.004)	0.062 (0.003)	0.179 (0.002)	0.872 (0.005)	1.5
JKSE	0.063 (0.005)	0.096 (0.002)	0.190 (0.001)	0.856 (0.005)	1.5
KLSE	0.087 (0.022)	0.071 (0.002)	0.157 (0.001)	0.835 (0.014)	2
N225	0.044 (0.004)	0.038 (0.003)	0.148 (0.002)	0.898 (0.006)	1
NZ50	0.018 (0.019)	0.044 (0.006)	0.120 (0.004)	0.898 (0.010)	1.5
STI	0.027 (0.011)	0.078 (0.001)	0.178 (0.001)	0.876 (0.005)	1.5
KS11	0.017 (0.009)	0.049 (0.001)	0.121 (0.004)	0.923 (0.008)	1.5
TWII	0.028 (0.012)	0.041 (0.004)	0.123 (0.003)	0.918 (0.010)	1
ATX	0.030 (0.005)	0.050 (0.002)	0.137 (0.003)	0.902 (0.007)	1
BFX	0.027 (0.005)	0.028 (0.002)	0.154 (0.003)	0.898 (0.005)	1.5
FCHI	0.026 (0.008)	0.014 (0.003)	0.112 (0.004)	0.931 (0.009)	1
GDAXI	0.028 (0.010)	0.022 (0.003)	0.114 (0.006)	0.926 (0.011)	1
AEX.AS	0.019 (0.005)	0.030 (0.002)	0.130 (0.002)	0.917 (0.005)	1.5
SSMI	0.038 (0.008)	0.024 (0.003)	0.145 (0.004)	0.897 (0.008)	1
FTSE	0.015 (0.010)	0.017 (0.003)	0.111 (0.003)	0.935 (0.008)	1
GD.AT	0.045 (0.001)	0.104 (0.002)	0.157 (0.001)	0.865 (0.004)	2
TA100	0.088 (0.007)	0.057 (0.002)	0.178 (0.001)	0.854 (0.007)	1.5

Table 3. Correlation matrix estimate $\hat{R}$

	MER	BVS	GST	MXX	GSC	AOR	SSE	HSI	BSE	JKS	KLS	N22	NZ5
MERV	1.00												
BVSP	0.53	1.00											
GSPT	0.47	0.48	1.00										
MXX	0.47	0.52	0.48	1.00									
GSPC	0.48	0.52	0.67	0.55	1.00								
AORD	0.17	0.17	0.21	0.17	0.12	1.00							
SSEC	0.06	0.08	0.08	0.06	0.02	0.18	1.00						
HSI	0.21	0.19	0.22	0.21	0.14	0.49	0.28	1.00					
BSES	0.17	0.19	0.21	0.20	0.15	0.31	0.14	0.40	1.00				
JKSE	0.15	0.15	0.14	0.15	0.08	0.36	0.15	0.43	0.31	1.00			
KLSE	0.10	0.10	0.11	0.12	0.06	0.28	0.14	0.36	0.19	0.32	1.00		
N225	0.11	0.13	0.19	0.12	0.12	0.46	0.16	0.44	0.27	0.34	0.28	1.00	
NZ50	0.09	0.06	0.10	0.09	0.04	0.48	0.16	0.31	0.21	0.29	0.22	0.38	1.00
STI	0.22	0.20	0.22	0.20	0.16	0.44	0.18	0.56	0.38	0.44	0.39	0.40	0.32
KS11	0.15	0.20	0.20	0.20	0.15	0.49	0.16	0.55	0.33	0.36	0.27	0.54	0.32
TWII	0.13	0.14	0.15	0.13	0.10	0.41	0.18	0.47	0.27	0.33	0.27	0.44	0.31
ATX	0.31	0.27	0.33	0.30	0.30	0.32	0.12	0.33	0.27	0.28	0.19	0.27	0.22
BFX	0.35	0.33	0.40	0.36	0.42	0.30	0.09	0.31	0.27	0.24	0.17	0.25	0.20
FCHI	0.37	0.36	0.44	0.39	0.47	0.26	0.06	0.31	0.28	0.21	0.15	0.26	0.17
GDAX	0.36	0.37	0.44	0.38	0.47	0.30	0.07	0.34	0.28	0.21	0.16	0.27	0.16
AEX	0.37	0.36	0.45	0.39	0.45	0.31	0.06	0.35	0.29	0.22	0.18	0.28	0.18
SSMI	0.33	0.31	0.39	0.35	0.41	0.29	0.05	0.31	0.27	0.23	0.16	0.27	0.19
FTSE	0.38	0.37	0.46	0.39	0.47	0.28	0.06	0.32	0.29	0.22	0.17	0.27	0.18
GD	0.19	0.18	0.20	0.19	0.16	0.21	0.07	0.24	0.26	0.20	0.14	0.19	0.17
TA10	0.24	0.24	0.27	0.26	0.23	0.33	0.06	0.36	0.28	0.24	0.18	0.29	0.18

	STI	KS1	TWI	ATX	BFX	FCH	GDA	AEX	SSM	FTS	GD	TA1
STI	1.00											
KS11	0.50	1.00										
TWII	0.45	0.51	1.00									
ATX	0.32	0.28	0.23	1.00								
BFX	0.30	0.25	0.19	0.56	1.00							
FCHI	0.30	0.26	0.20	0.55	0.71	1.00						
GDAX	0.31	0.27	0.20	0.59	0.70	0.79	1.00					
AEX	0.33	0.28	0.22	0.58	0.74	0.82	0.79	1.00				
SSMI	0.30	0.26	0.21	0.52	0.66	0.72	0.72	0.74	1.00			
FTSE	0.31	0.27	0.19	0.54	0.66	0.77	0.70	0.76	0.69	1.00		
GD	0.25	0.27	0.21	0.32	0.34	0.34	0.33	0.33	0.32	0.30	1.00	
TA10	0.36	0.28	0.25	0.38	0.39	0.42	0.40	0.41	0.40	0.40	0.33	1.00

Table 4. Correlations between the variables and the first 3 factors of the PCA

	PC1	PC2	PC3		PC1	PC2	PC3
MER	-0.52	-0.29	-0.46	STI	-0.58	0.45	-0.09
BVS	-0.52	-0.29	-0.52	KS1	-0.55	0.50	-0.11
GSPT	-0.59	-0.32	-0.41	TWI	-0.46	0.50	-0.11
MXX	-0.54	-0.30	-0.46	ATX	-0.68	-0.08	0.22
GSPC	-0.56	-0.45	-0.41	BFX	-0.75	-0.25	0.27
AOR	-0.55	0.46	-0.02	FCH	-0.79	-0.32	0.28
SSE	-0.19	0.27	-0.14	GDA	-0.79	-0.29	0.27
HSI	-0.60	0.48	-0.07	AEX	-0.81	-0.28	0.29
BSE	-0.48	0.25	-0.04	SSM	-0.75	-0.24	0.30
JKS	-0.45	0.42	-0.06	FTS	-0.78	-0.28	0.21
KLS	-0.35	0.38	-0.07	GD.	-0.46	0.05	0.14
N22	-0.50	0.47	-0.00	TA1	-0.57	0.06	0.10
NZ5	-0.37	0.44	0.03				

and cover the period from January 14, 2000 to May 16, 2013, which corresponds to 2081 observations. On these 6 series, we fitted an extended CCC-GARCH(1,1) model of the form

$$\underline{h}_t = \underline{\omega} + \mathbf{A}\underline{\epsilon}_{t-1} + \mathbf{B}\underline{h}_{t-1}$$

where $\mathbf{B}$ is diagonal. This assumption allows to fit the model equation by equation. The estimated values of $\mathbf{A}$ and $\mathbf{B}$ are

$$\hat{\mathbf{A}} = \begin{pmatrix} \underset{0.010}{0.029} & \underset{0.003}{0.002} & \underset{0.040}{0.015} & \underset{0.013}{0.012} & \underset{0.003}{0.003} & \underset{0.038}{0.000} \\ \underset{0.002}{0.000} & \underset{0.023}{0.136} & \underset{0.004}{0.000} & \underset{0.003}{0.003} & \underset{0.001}{0.000} & \underset{0.003}{0.000} \\ \underset{0.005}{0.000} & \underset{0.002}{0.002} & \underset{0.028}{0.031} & \underset{0.007}{0.008} & \underset{0.002}{0.002} & \underset{0.027}{0.001} \\ \underset{0.004}{0.006} & \underset{0.002}{0.001} & \underset{0.020}{0.004} & \underset{0.012}{0.041} & \underset{0.002}{0.006} & \underset{0.019}{0.000} \\ \underset{0.012}{0.017} & \underset{0.005}{0.003} & \underset{0.054}{0.000} & \underset{0.016}{0.002} & \underset{0.012}{0.061} & \underset{0.052}{0.000} \\ \underset{0.005}{0.000} & \underset{0.002}{0.003} & \underset{0.028}{0.024} & \underset{0.007}{0.007} & \underset{0.002}{0.002} & \underset{0.028}{0.008} \end{pmatrix}, \quad \hat{\mathbf{B}} = \mathrm{diag} \begin{pmatrix} \underset{0.022}{0.92} \\ \underset{0.017}{0.88} \\ \underset{0.010}{0.95} \\ \underset{0.015}{0.93} \\ \underset{0.014}{0.93} \\ \underset{0.009}{0.96} \end{pmatrix},$$

and the estimation of the correlation matrix $\boldsymbol{R}$ is

$$\hat{\boldsymbol{R}} = \begin{pmatrix} 1.00 & \underset{\scriptscriptstyle 0.026}{0.00} & \underset{\scriptscriptstyle 0.039}{0.46} & \underset{\scriptscriptstyle 0.031}{0.39} & \underset{\scriptscriptstyle 0.034}{0.17} & \underset{\scriptscriptstyle 0.032}{0.47} \\ 0.00 & 1.00 & \underset{\scriptscriptstyle 0.040}{0.14} & \underset{\scriptscriptstyle 0.027}{0.12} & \underset{\scriptscriptstyle 0.043}{0.42} & \underset{\scriptscriptstyle 0.045}{0.13} \\ 0.46 & 0.14 & 1.00 & \underset{\scriptscriptstyle 0.033}{0.44} & \underset{\scriptscriptstyle 0.039}{0.58} & \underset{\scriptscriptstyle 0.031}{0.98} \\ 0.39 & 0.12 & 0.44 & 1.00 & \underset{\scriptscriptstyle 0.071}{0.26} & \underset{\scriptscriptstyle 0.040}{0.45} \\ 0.17 & 0.42 & 0.58 & 0.26 & 1.00 & \underset{\scriptscriptstyle 0.044}{0.57} \\ 0.47 & 0.13 & 0.98 & 0.45 & 0.57 & 1.00 \end{pmatrix} \begin{matrix} \text{CAD} \\ \text{CHF} \\ \text{CNY} \\ \text{GBP} \\ \text{JPY} \\ \text{USD} \end{matrix}$$

The estimated standard deviations of the estimators were obtained from Theorem 4.2 and are displayed in small font size. It can be noted that the volatilities of the different exchange rates are mainly linked by the strong correlations of the residuals, which can be interpreted as a sign of instantaneous causality between the squared returns. By contrast, in view of the diagonal form of $\hat{\mathbf{A}}$, the volatility of a given exchange rate is mainly explained by its own past returns. A noticeable exception is the volatility of the USD which shows more sensitivity to the variations of the CNY than to its own variations. These two exchange rates are also strongly related by the correlation (0.98) between their rescaled residuals.

We now relax the constant correlation assumption (2.7) by considering a DCC matrix $\boldsymbol{R}_t^*$ of the form (4.2) with $N = 2$ regimes. The estimates of the GARCH(1,1) parameters are unchanged, but the estimated CCC matrix $\hat{\boldsymbol{R}}$ is replaced by the following estimates of the correlation matrix in each of the two regimes

$$\hat{\boldsymbol{R}}^*(1) = \begin{pmatrix} 1.00 & \underset{\scriptscriptstyle 0.150}{0.38} & \underset{\scriptscriptstyle 0.062}{0.71} & \underset{\scriptscriptstyle 0.141}{0.69} & \underset{\scriptscriptstyle 0.127}{0.58} & \underset{\scriptscriptstyle 0.061}{0.72} \\ 0.38 & 1.00 & \underset{\scriptscriptstyle 0.138}{0.59} & \underset{\scriptscriptstyle 0.107}{0.52} & \underset{\scriptscriptstyle 0.066}{0.66} & \underset{\scriptscriptstyle 0.140}{0.59} \\ 0.71 & 0.59 & 1.00 & \underset{\scriptscriptstyle 0.132}{0.81} & \underset{\scriptscriptstyle 0.096}{0.89} & \underset{\scriptscriptstyle 0.002}{0.99} \\ 0.69 & 0.52 & 0.81 & 1.00 & \underset{\scriptscriptstyle 0.146}{0.76} & \underset{\scriptscriptstyle 0.135}{0.82} \\ 0.58 & 0.66 & 0.89 & 0.76 & 1.00 & \underset{\scriptscriptstyle 0.101}{0.90} \\ 0.72 & 0.59 & 0.99 & 0.82 & 0.90 & 1.00 \end{pmatrix}$$

and

$$\hat{\boldsymbol{R}}^*(2) = \begin{pmatrix} 1.00 & \underset{0.039}{-0.04} & \underset{0.029}{0.42} & \underset{0.030}{0.34} & \underset{0.042}{0.10} & \underset{0.028}{0.43} \\ -0.04 & 1.00 & \underset{0.044}{0.08} & \underset{0.039}{0.08} & \underset{0.028}{0.39} & \underset{0.044}{0.07} \\ 0.42 & 0.08 & 1.00 & \underset{0.039}{0.38} & \underset{0.033}{0.52} & \underset{0.001}{0.98} \\ 0.34 & 0.08 & 0.38 & 1.00 & \underset{0.051}{0.18} & \underset{0.039}{0.38} \\ 0.10 & 0.39 & 0.52 & 0.18 & 1.00 & \underset{0.034}{0.51} \\ 0.43 & 0.07 & 0.98 & 0.38 & 0.51 & 1.00 \end{pmatrix}.$$

The estimated standard deviations of the estimators, displayed in small font size, are obtained by taking the empirical standard deviations of the estimates of $N = 100$ independent simulations of the DCC model that have been fitted on the real data set.

The transition probabilities of the Markov chain are estimated by $\hat{p}(1,1) = 0.826$, $\hat{p}(1,2) = 0.174$, $\hat{p}(2,1) = 0.039$ and $\hat{p}(2,2) = 0.961$, with respective estimated standard deviations 0.036, 0.036, 0.013 and 0.013. This corresponds to regimes with relative frequencies $\hat{P}(\Delta_t = 1) = 0.18$ and $\hat{P}(\Delta_t = 2) = 0.82$. The second regime being the most frequent, it is not surprising to observe that $\hat{\boldsymbol{R}}^*(2)$ and $\hat{\boldsymbol{R}}$ are close. It seems however that the introduction of two regimes is relevant. Indeed, the less frequent regime is characterized by significantly more correlated residuals. Figure 2 illustrates the high positive correlation between the GBP and JPY residuals when the most probable regime is the first one (left figure). Figure 3 shows that the regime with the highest residual correlations (*i.e.* the regime 1) is often more plausible when the volatilities are high.

Finally, we tested the adequacy of BEKK models, using the results of Section 5. For each pair of exchange rates, we estimated Model (5.2) and we tested the restrictions $H_0^{(1)}$ and $H_0^{(2)}$ that are satisfied when the DGP is the BEKK-GARCH(1,1) model (5.1). Table 5 shows that, for 12 bivariate series over 15, either $H_0^{(1)}$ or $H_0^{(2)}$ is clearly rejected, which invalidates the adequacy of this BEKK model for the 12 series. Using the Bonferroni correction, one can indeed reject the model at the significant level less than α if one of the two hypothesis $H_0^{(k)}$ is rejected at the level $\alpha/2$.

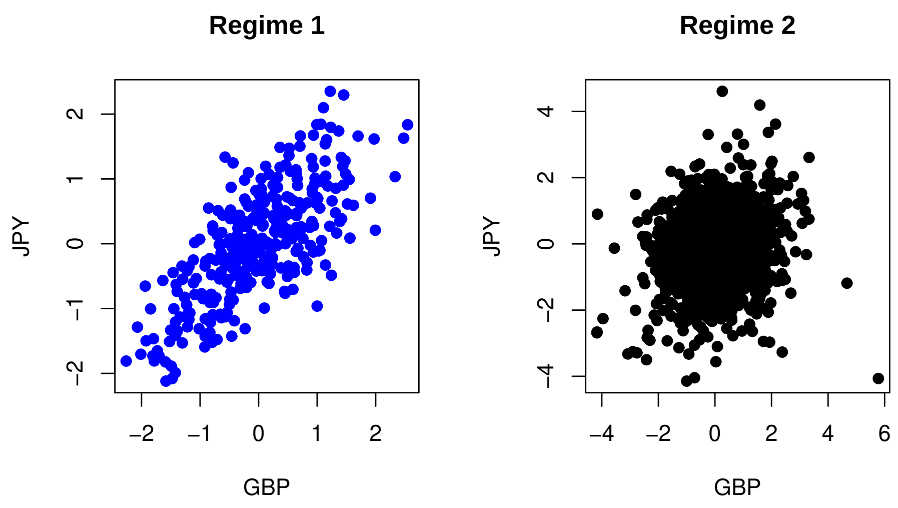

Figure 2. GBP and JPY residuals as function of the most probable regime

Table 5. For each pair of exchange rates: p-values of the tests of the null hypotheses $H_0^{(1)}$ and $H_0^{(2)}$ implied by the bivariate BEKK-GARCH(1,1) model. Gray cells contain p-values less than 2.5%.

	CAD		CHF		CNY		GBP		JPY	
	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$
CHF	0.000	0.163								
CNY	0.120	0.015	0.122	0.500						
GBP	0.012	0.023	0.128	0.000	0.005	0.100				
JPY	0.007	0.006	0.500	0.500	0.500	0.087	0.050	0.000		
USD	0.500	0.021	0.114	0.000	0.500	0.381	0.068	0.000	0.102	0.000

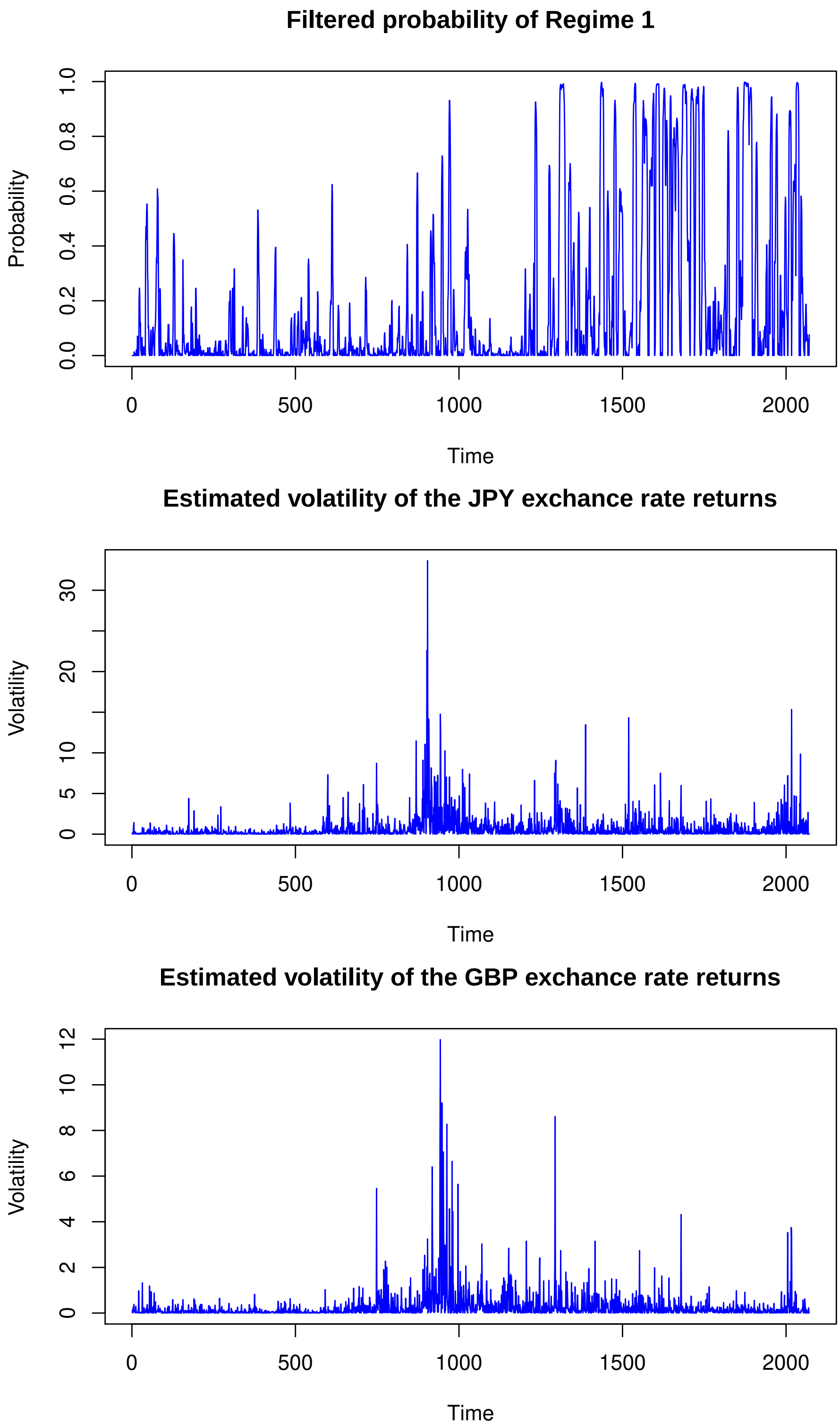

Figure 3. Filtered probability of Regime 1, and estimated volatilities of the GBP and JPY exchange rate returns

7. Conclusion

We studied a method allowing for simple estimation of MGARCH and SC models. Instead of applying the full QML to the whole set of parameters, which is often numerically inefficient (see Table 1) if not infeasible, we estimated the volatility parameters by QML equation-by-equation in a first step, and then used the volatility-standardized returns in a second step to estimate the conditional correlation matrix. In contrast to other methods which have been proposed in the literature, this approach does not make strong a priori restrictions on the volatilities of the individual returns, which may be general functions of the past values of all returns.

Our aim was not only to develop an easily implementable MGARCH estimation procedure, but also to derive asymptotic estimation results under mild assumptions on the observed process. The complexity of MGARCH specifications often make the asymptotic properties of the QMLE difficult to establish. By contrast, the simplicity of the proposed procedure allows for a rigorous analysis of asymptotic theory.

Moreover, our procedure is compatible with different assumptions on the conditional correlation matrix. The constant case was studied in details, and we obtained the joint asymptotic distribution of the volatility and correlation parameters. Such results are amenable to different extensions, one of which was considered in the paper (*i.e.* the hidden Markov model for the correlation matrix). We also used our first step estimator to test the BEKK specification. Other extensions are left for future research.

Appendix

A. Technical assumptions

We make the following assumptions on the volatility function.

A2: for any real sequence $(e_i)_{i\geq 1}$, the function $\boldsymbol{\theta}^{(k)} \mapsto \sigma_k(e_1, e_2, \dots; \boldsymbol{\theta}^{(k)})$ is continuous and there exists a measurable function $K : \mathbb{R}^\infty \mapsto (0, \infty)$ such that

$$|\sigma_k(e_1, e_2, \dots; \boldsymbol{\theta}^{(k)}) - \sigma_k(e_1, e_2, \dots; \boldsymbol{\theta}_0^{(k)})| \leq K(e_1, \dots)\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}\|,$$

and

$$E\left(\frac{K(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\right)^2 < \infty.$$

A3: there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left(\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right)^2 < \infty.$$

A4: we have $\sigma_{kt}(\cdot) > \underline{\omega}$ for some $\underline{\omega} > 0$.

A5: we have $\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) = \sigma_{kt}(\boldsymbol{\theta}^{(k)})$ $a.s.$ iff $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}_0^{(k)}$.

The next assumption allows to show that initial values have no effect on the asymptotic properties of the estimator of $\boldsymbol{\theta}_0^{(k)}$. Let $\Delta_{kt}(\boldsymbol{\theta}^{(k)}) = \tilde{\sigma}_{kt}(\boldsymbol{\theta}^{(k)}) - \sigma_{kt}(\boldsymbol{\theta}^{(k)})$, $a_t = \sup_k \sup_{\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}} |\Delta_{kt}(\boldsymbol{\theta}^{(k)})|$. Let C and ρ be generic constants with $C > 0$ and $0 < \rho < 1$. The "constant" C is allowed to depend on variables anterior to $t = 0$.

A6: We have $a_t \leq C\rho^t$, a.s.

To derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$, the following additional assumptions are considered.

A9: for any real sequence $(e_i)_{i \geq 1}$, the function $\boldsymbol{\theta}^{(k)} \mapsto \sigma_k(e_1, e_2, \ldots; \boldsymbol{\theta}^{(k)})$ has continuous second-order derivatives;

A10: there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\|^{4(1+\frac{1}{\delta})}, \quad \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial^2 \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} \right\|^{2(1+\frac{1}{\delta})},$$

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left| \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right|^4,$$

have finite expectations.

The next assumption is introduced to handle initial values.

A11: We have

$$b_t := \sup_k \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial \Delta_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\| \leq C\rho^t, \quad a.s.$$

The next assumption will be used to show the invertibility of the asymptotic covariance matrix.

A12: For $k = 1, \dots, m$ and for any $\boldsymbol{x} \in \mathbb{R}^{d_k}$, we have:

$$\boldsymbol{x}' \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0, \ a.s. \quad \Rightarrow \quad \boldsymbol{x} = 0.$$

The next assumption is used in Theorem 3.2.

A10*: there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\|^4, \qquad \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial^2 \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\|^2,$$

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left| \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right|^4,$$

have finite expectations.

B. Proofs

B.1. Proof of Theorem 3.1

a) The strong consistency of $\hat{\boldsymbol{\theta}}_n^{(k)}$ is a consequence of the following intermediate results:

$$i) \lim_{n\to\infty} \sup_{\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}} |Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) - \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)})| = 0, \quad a.s.,$$

$$ii) \ \mathbb{E}|\ell_{k,1}(\boldsymbol{\theta}_0^{(k)})| < \infty, \quad \text{and if } \boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}, \quad \mathbb{E}\ell_{k,1}(\boldsymbol{\theta}_0^{(k)}) < \mathbb{E}\ell_{k,1}(\boldsymbol{\theta}^{(k)}),$$

$iii)$ any $\boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}$ has a neighborhood $V(\boldsymbol{\theta}^{(k)})$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta}^{(k)})} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^*) > \limsup_{n\to\infty} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)}), \ a.s.$$

Because the proof follows along the same lines as the proof of Theorem 7.1 in Francq and Zakoïan (2010) we omit details. It is easy to see that i) follows from **A4**, **A6** and the existence of $E|\epsilon_{kt}|^s$. To show ii), first note that $E\left(\eta_{kt}^* \mid \mathcal{F}_{t-1}^{\epsilon}\right) = 0$. In view of (2.3), we thus have

$$E\ell_{kt}(\boldsymbol{\theta}^{(k)}) = E\left\{ \frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})\eta_{kt}^{*2}}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} + \log \sigma_{kt}^2(\boldsymbol{\theta}^{(k)}) \right\} = E\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} + E\log \sigma_{kt}^2(\boldsymbol{\theta}^{(k)}) < \infty.$$

Since $E\log\sigma_{kt}^2 < \infty$, we have $E\ell_{kt}(\boldsymbol{\theta}_0^{(k)}) < \infty$, whereas $E\ell_{kt}(\boldsymbol{\theta}^{(k)}) > -\infty$, for any $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$, by **A4**. Using the elementary inequality $\log x \leq x - 1$ and **A5**, ii) follows. The last point follows from the ergodic theorem, which can be applied for any $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$ to the sequence $\inf_{\boldsymbol{\theta}_*\in V(\boldsymbol{\theta}^{(k)})\cap\boldsymbol{\theta}^{(k)}} \ell_{kt}(\boldsymbol{\theta}_*)$, which is strictly stationary and ergodic under **A1** and admits an expectation in $(-\infty, \infty]$.

b) Now we turn to the proof of the asymptotic normality. Define $\tilde{\ell}_{kt}$ as ℓ_{kt}, with σ_{kt} replaced by $\tilde{\sigma}_{kt}$. The proof relies on a set of preliminary results.

$$i)\; E\left\|\frac{\partial \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}\frac{\partial \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)\prime}}\right\| < \infty, \quad E\left\|\frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}}\right\| < \infty,$$

$ii)$ There exists a neighbourhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} - \frac{\partial \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}\right\| \to 0\,,$$

$$iii)\; \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_n^{(k)})}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}} \to \boldsymbol{J}_{kk}\,, \quad \text{a.s.} \quad \text{for any } \boldsymbol{\theta}_n^{(k)} \text{ between } \hat{\boldsymbol{\theta}}_n^{(k)} \text{ and } \boldsymbol{\theta}_0^{(k)},$$

$iv)$ $\boldsymbol{J}_{kk}$ is non singular,

$$v)\; \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}_{kk}\right).$$

Note that

$$\begin{aligned}
\frac{\partial \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} &= \left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\},\\
\frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}} &= \left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial^2 \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}}\right\}\\
&\quad +2\left\{3\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}-1\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)\prime}}\right\}.
\end{aligned}$$

Let $\|\cdot\|_r$ denote the L^r norm, for $r \geq 1$, on the space of real random variables. We have, by the Hölder inequality,

$$\left\|\left(1-\eta_{kt}^{*2}\right)\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}(\boldsymbol{\theta}_0^{(k)})\right\|_2 \leq \|1-\eta_{kt}^{*2}\|_{2(\delta+1)}\left\|\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}(\boldsymbol{\theta}_0^{(k)})\right\|_{2(1+1/\delta)},$$

which is finite by Assumptions **A8** and **A10**. The first result in i) follows. The second result can be shown similarly.

Now, turning to ii), we have

$$\begin{aligned}
&\left\|\frac{\partial \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} - \frac{\partial \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}\right\|\\
=\;&\left\|\left\{\frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2}-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}+2\left\{1-\frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2}\right\}\left\{\frac{1}{\sigma_{kt}}-\frac{1}{\tilde{\sigma}_{kt}}\right\}\left\{\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}\right.\\
&\left.+\left\{1-\frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2}\right\}\left\{\frac{2}{\tilde{\sigma}_{kt}}\right\}\left\{\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}-\frac{\partial \tilde{\sigma}_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}\right\|(\boldsymbol{\theta}^{(k)}) \leq C\rho^t u_t,
\end{aligned}$$

where

$$u_t = (1+\eta_{kt}^{*2})\left(1+\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sigma_{kt}}\frac{\partial\sigma_{kt}}{\partial\boldsymbol{\theta}^{(k)}}(\boldsymbol{\theta}^{(k)})\right\|\right)\left(1+\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left|\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\right|^2\right),$$

as a consequence of Assumptions **A4, A6** and **A11**. We have $E|u_t|<\infty$ by Assumptions **A8** and **A10**, and using the Cauchy-Schwarz inequality. Thus $C\sum_{t=1}^n \rho^t u_t$ is bounded a.s., which entails ii).

To prove iii), by Exercise 7.9 in Francq and Zakoïan (2010), it will be sufficient to establish that for any $\varepsilon>0$, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n \sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{\partial^2\ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}-\frac{\partial^2\ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}\right\|\le\varepsilon \quad a.s. \tag{B.2}$$

By the ergodic theorem, the limit in the left-hand side is equal to

$$E\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{\partial^2\ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}-\frac{\partial^2\ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}\right\|$$

provided that this expectation is finite. In view of **A9**, the conclusion will follow by the dominated convergence theorem: the latter expectation tends to zero when the neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ shrinks to the singleton $\{\boldsymbol{\theta}_0^{(k)}\}$. To complete the proof of iii), it thus remains to show that

$$E\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{\partial^2\ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}\right\|<\infty. \tag{B.3}$$

Let us consider the first product in the right-hand side of (B.1)). We have, by the Hölder inequality,

$$\begin{aligned}&E\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial^2\sigma_{kt}}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}\right\}\right\|\\ \le\ &\left\{1+\|\eta_{kt}^{*2}\|_{2(1+\delta)}\left\|\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}\right\|_2\right\}\left\|\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sigma_{kt}}\frac{\partial^2\sigma_{kt}}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}(\boldsymbol{\theta}^{(k)})\right\|\right\|_{2(1+1/\delta)},\end{aligned}$$

which is finite by Assumptions **A8** and **A10**. The second product in the right-hand side of (B.1)) can be handled similarly. Thus iii) is established.

The invertibility of $\boldsymbol{J}_{kk}$ is a straightforward consequence of **A12**. Now

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial\ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}^{(k)}}=\frac{1}{\sqrt{n}}\sum_{t=1}^n\{1-\eta_{kt}^{*2}\}\boldsymbol{d}_{kt},$$

and v) follows from the Central Limit Theorem of Billingsley (1961) for ergodic, stationary and square integrable martingale differences. Indeed, the square integrability follows again from the Cauchy-Schwarz inequality,

$$E\left(\{1-\eta_{kt}^{*2}\}^2\|\boldsymbol{d}_{kt}\boldsymbol{d}_{kt}'\|\right) \leq \|(1-\eta_{kt}^{*2})^2\|_{1+\delta}\|\boldsymbol{d}_{kt}\boldsymbol{d}_{kt}'\|_{1+1/\delta},$$

and Assumptions **A8** and **A10**. Moreover, $(\boldsymbol{\eta}_t^*)$ is strictly stationary and ergodic as a function of the process $(\boldsymbol{\epsilon}_t)$.

We are now in a position to complete the proof of Theorem 3.1. Since $\hat{\boldsymbol{\theta}}_n^{(k)}$ converges to $\boldsymbol{\theta}_0^{(k)}$, which stands in the interior of the parameter space by **A7**, the derivative of the criterion $\tilde{Q}_n^{(k)}$ is equal to zero at $\hat{\boldsymbol{\theta}}_n^{(k)}$. In view of point ii), we thus have by a Taylor expansion of $Q_n^{(k)}$ at $\boldsymbol{\theta}_0^{(k)}$,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)}-\boldsymbol{\theta}_0^{(k)}\right) \overset{o_P(1)}{=} -\left(\frac{1}{n}\sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_{ij}^{*(k)})}{\partial\theta_i^{(k)}\partial\theta_j^{(k)}}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial}{\partial\boldsymbol{\theta}^{(k)}}\ell_{kt}(\boldsymbol{\theta}_0^{(k)})$$

where the $\boldsymbol{\theta}_{ij}^{*(k)}$'s are between $\hat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$. The conclusion follows from the intermediate results i)-v). □

B.2. Proof of the result of Section 3.2

We have

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta}^{(1)})}{\partial\boldsymbol{\theta}^{(1)}} &= \left\{1-\frac{1}{1-\rho_0^2}\left(\frac{\epsilon_{1t}}{\sigma_{1t}}-\rho_0\frac{\epsilon_{2t}}{\sigma_{2t}}\right)\frac{\epsilon_{1t}}{\sigma_{1t}}\right\}\left\{\frac{2}{\sigma_{1t}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\right\},\\
\frac{\partial^2 l_t(\boldsymbol{\theta}^{(1)})}{\partial\boldsymbol{\theta}^{(1)}\partial\boldsymbol{\theta}^{(1)\prime}} &= -\frac{1}{1-\rho_0^2}\left\{-2\frac{\epsilon_{1t}}{\sigma_{1t}}+\rho_0\frac{\epsilon_{2t}}{\sigma_{2t}}\right\}\frac{\epsilon_{1t}}{\sigma_{1t}}\left\{\frac{2}{\sigma_{1t}^2}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)\prime}}\right\},\\
&\quad+\left\{1-\frac{1}{1-\rho_0^2}\left(\frac{\epsilon_{1t}}{\sigma_{1t}}-\rho_0\frac{\epsilon_{2t}}{\sigma_{2t}}\right)\frac{\epsilon_{1t}}{\sigma_{1t}}\right\}\frac{\partial}{\partial\boldsymbol{\theta}^{(1)\prime}}\left\{\frac{2}{\sigma_{1t}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\right\}.
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta}_0^{(1)})}{\partial\boldsymbol{\theta}^{(1)}} &= \left\{1-\frac{1}{1-\rho_0^2}\left(\eta_{1t}^*-\rho_0\eta_{2t}^*\right)\eta_{1t}^*\right\}\left\{\frac{2}{\sigma_{1t}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\right\},\\
\frac{\partial^2 l_t(\boldsymbol{\theta}_0^{(1)})}{\partial\boldsymbol{\theta}^{(1)}\partial\boldsymbol{\theta}^{(1)\prime}} &= -\frac{1}{1-\rho_0^2}\left\{-2\eta_{1t}^*+\rho_0\eta_{2t}^*\right\}\eta_{1t}^*\left\{\frac{2}{\sigma_{1t}^2}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)\prime}}\right\}.
\end{aligned}$$

We thus have

$$\begin{aligned}
\mathrm{Var}\left\{\frac{\partial l_t(\boldsymbol{\theta}_0^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right\} &= \zeta E\left\{\frac{4}{\sigma_{1t}^2}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)\prime}}(\boldsymbol{\theta}_0^{(1)})\right\},\\
E\left(\frac{\partial^2 l_t(\boldsymbol{\theta}_0^{(1)})}{\partial\boldsymbol{\theta}^{(1)}\partial\boldsymbol{\theta}^{(1)\prime}}\right) &= \frac{2-\rho_0^2}{1-\rho_0^2}E\left\{\frac{2}{\sigma_{1t}^2}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)\prime}}(\boldsymbol{\theta}_0^{(1)})\right\}.
\end{aligned}$$

Under appropriate conditions, the asymptotic variance of the FQMLE is given by

$$\begin{aligned}\mathbf{\Sigma}^{FQML} &= \left\{E\left(\frac{\partial^2 l_t(\boldsymbol{\theta}^{(1)})}{\partial\boldsymbol{\theta}^{(1)\prime}\partial\boldsymbol{\theta}^{(1)}}\right)\right\}^{-1}\mathrm{Var}\left\{\frac{\partial l_t(\boldsymbol{\theta}_0^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right\}\left\{E\left(\frac{\partial^2 l_t(\boldsymbol{\theta}^{(1)})}{\partial\boldsymbol{\theta}^{(1)\prime}\partial\boldsymbol{\theta}^{(1)}}\right)\right\}^{-1} \\ &= \left(\frac{1-\rho_0^2}{2-\rho_0^2}\right)^2\zeta\left\{E\left(\frac{1}{\sigma_{1t}^2}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)\prime}}(\boldsymbol{\theta}_0^{(1)})\right)\right\}^{-1}.\end{aligned}$$

On the other hand, the asymptotic variance of the QMLE of $\boldsymbol{\vartheta}^{(1)}$ based on the single equation of ϵ_{1t} is

$$\mathbf{\Sigma}_1 = (E\eta_{1t}^{*4}-1)\left\{E\left(\frac{4}{\sigma_{1t}^2}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)}}\frac{\partial\sigma_{1t}}{\partial\boldsymbol{\theta}^{(1)\prime}}(\boldsymbol{\theta}_0^{(1)})\right)\right\}^{-1}.$$

The conclusion follows. □

B.3. Proof of Theorem 3.2

Note that under the independence assumption (3.4),

$$E|\eta_{kt}^*|^s < \infty \quad \text{and} \quad E|\epsilon_{kt}|^s = E|\sigma_{kt}|^s E|\eta_{kt}^*|^s < \infty$$

imply $E|\sigma_{kt}|^s < \infty$. Therefore the condition $E\log\sigma_{kt}^2 < \infty$ can be omitted in **A1**.

The proof of the asymptotic normality relies on the same steps *i)-v)* as in the proof of the second part of Theorem 3.1, except that one can replace $\boldsymbol{I}_{kk}$ by $(E\eta_{k1}^4-1)\boldsymbol{J}_{kk}$, and the assumptions **A8** and **A10** by **A8*** and **A10***. In particular, to show (B.3), note that, by the Cauchy-Schwarz inequality,

$$\begin{aligned}& E\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial^2\sigma_{kt}}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}\right\}\right\| \\ \leq & \left\{1+\|\eta_{kt}^*\|_2\left\|\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}\right\|_2\right\}\left\|\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sigma_{kt}}\frac{\partial^2\sigma_{kt}}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}(\boldsymbol{\theta}^{(k)})\right\|\right\|_2,\end{aligned}$$

which is finite under Assumptions **A8*** and **A10***. □

B.4. Proof of Proposition 3.1

Recall that for any spherically distributed variable $\boldsymbol{X} = (X_1,\dots,X_m)'$, we have $\boldsymbol{\lambda}'\boldsymbol{X} \stackrel{d}{=} \|\boldsymbol{\lambda}\|X_1$ for any $\boldsymbol{\lambda}\in\mathbb{R}^m$, where $\stackrel{d}{=}$ stands for equality in distribution and $\|\cdot\|$ denotes the Euclidian norm on $\mathbb{R}^m$. Letting $\boldsymbol{e}_k$ the k-th column of $\boldsymbol{I}_m$, we have

$$\eta_{kt}^* = \boldsymbol{e}_k'\boldsymbol{R}_t^{*1/2}\boldsymbol{\xi}_t \stackrel{d}{=} \|\boldsymbol{e}_k'\boldsymbol{R}_t^{*1/2}\|\xi_1 = \xi_1 \qquad \text{(B.4)}$$

conditionally to $\boldsymbol{R}_t^*$, and thus unconditionally.

Now for any $x, y \in \mathbb{R}$, using successively the independence between $\boldsymbol{\xi}_t$ et $\boldsymbol{\xi}_{t-1}$ and the independence between $(\boldsymbol{R}_t^*)$ and $(\boldsymbol{\xi}_t)$, for $k, \ell = 1, \dots, m$,

$$\begin{aligned} P(\eta_{kt}^* < x, \eta_{\ell,t-1}^* < y \mid \boldsymbol{R}_t^*, \boldsymbol{R}_{t-1}^*) &= P(\eta_{kt}^* < x \mid \boldsymbol{R}_t^*, \boldsymbol{R}_{t-1}^*) P(\eta_{\ell,t-1}^* < y \mid \boldsymbol{R}_t^*, \boldsymbol{R}_{t-1}^*) \\ &= P(\eta_{kt}^* < x \mid \boldsymbol{R}_t^*) P(\eta_{\ell,t-1}^* < y \mid \boldsymbol{R}_{t-1}^*) \\ &= P(\eta_{kt}^* < x) P(\eta_{\ell,t-1}^* < y), \end{aligned}$$

the last equality following from (B.4). We similarly prove that for any positive integer j

$$P(\eta_{k_1 t}^* < x_1, \dots, \eta_{k_j, t-j+1}^* < x_j) = \prod_{i=1}^{j} P(\eta_{k_i, t-i+1}^* < x_i)$$

for all sequences (k_i) and (x_i). The conclusion follows. □

B.5. Proof of Theorem 4.1

The consistency of $\hat{\boldsymbol{\theta}}_n$ follows from Theorem 3.1. It suffices to prove the consistency of $\hat{\boldsymbol{\rho}}_n$.

Let vec denote the operator that stacks the columns of a matrix. Let $\boldsymbol{K}_m$ denote a $m(m-1)/2 \times m^2$ matrix such that for any symmetric $m \times m$ matrix $\boldsymbol{A}$, $\boldsymbol{K}_m \mathrm{vec}(\boldsymbol{A}) = \mathrm{vech}^0(\boldsymbol{A})$. We have

$$\hat{\boldsymbol{\rho}}_n = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{K}_m \left(\hat{\boldsymbol{\eta}}_t^* \otimes \hat{\boldsymbol{\eta}}_t^* \right).$$

Letting

$$\boldsymbol{\rho}_n = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{K}_m \left(\boldsymbol{\eta}_t^* \otimes \boldsymbol{\eta}_t^* \right),$$

we have

$$\|\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n\| \leq \frac{C}{n} \sum_{t=1}^{n} \|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\| (\|\boldsymbol{\eta}_t^*\| + \|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\|).$$

Now, using **A2** and **A4**,

$$\begin{aligned}
\|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\| &\leq C\sum_{k=1}^m \frac{|\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) - \tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})|}{\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}|\eta_{kt}^*| \\
&\leq C\sum_{k=1}^m \frac{|\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) - \sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})| + |\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)}) - \tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})|}{\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}|\eta_{kt}^*| \\
&\leq C\sum_{k=1}^m \left(\frac{|\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) - \sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})|}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}\frac{\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}{\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})} + a_t\right)|\eta_{kt}^*| \\
&\leq C\sum_{k=1}^m \left(\frac{K(\boldsymbol{\epsilon}_{t-1},\dots)\|\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\|}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}(1+a_t) + a_t\right)|\eta_{kt}^*|.
\end{aligned}$$

We thus have, by **A6**, for n large enough such that $\hat{\boldsymbol{\theta}}_n^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})$,

$$\begin{aligned}
\|\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n\| &\leq \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|\frac{C}{n}\sum_{t=1}^n \|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m \frac{K(\boldsymbol{\epsilon}_{t-1},\dots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \\
&\quad + \frac{C}{n}\sum_{t=1}^n \rho^t\|\boldsymbol{\eta}_t^*\|^2 + \frac{C}{n}\sum_{t=1}^n \|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\|^2 := S_{n1} + S_{n2} + S_{n3}.
\end{aligned}$$

We have, using again the independence between $\boldsymbol{\eta}_t^*$ and $\{\boldsymbol{\epsilon}_u, u<t\}$ under (2.7),

$$\begin{aligned}
&E\left(\|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m \frac{K(\boldsymbol{\epsilon}_{t-1},\dots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\right) \\
&= E\|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m E\left(\frac{K(\boldsymbol{\epsilon}_{t-1},\dots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\right) \\
&\leq E\|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m \left\|\frac{K(\boldsymbol{\epsilon}_{t-1},\dots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\right\|_2 \left\|\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\right\|_2 < \infty,
\end{aligned}$$

using the Cauchy-Schwarz inequality. The last inequality is a consequence of Assumptions **A2-A3**. It follows that S_{n1} is the product of $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$ which converges to zero a.s., by Theorem 3.1, and a term which is bounded a.s. by the ergodic theorem. Thus $S_{n1} \to 0$ a.s. We similarly show that $S_{n2} \to 0$ and $S_{n3} \to 0$ a.s. Because $\boldsymbol{\eta}_t^* = \boldsymbol{R}^{1/2}\boldsymbol{\eta}_t$, the sequence $(\boldsymbol{\eta}_t^*)$ is iid . We thus have $\boldsymbol{\rho}_n \to \boldsymbol{\rho}_0$ by the strong law of large numbers. □

B.6. Proof of Theorem 4.2

Let

$$\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial\boldsymbol{\theta}^{(1)\prime}}\ell_{1t}(\boldsymbol{\theta}^{(1)}),\dots,\frac{\partial}{\partial\boldsymbol{\theta}^{(m)\prime}}\ell_{mt}(\boldsymbol{\theta}^{(m)})\right)'.$$

For $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$ let $\tilde{\eta}^*_{kt}(\boldsymbol{\theta}^{(k)}) = \tilde{\sigma}^{-1}_{kt}(\boldsymbol{\theta}^{(k)})\epsilon_{kt}$ and $\eta^*_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma^{-1}_{kt}(\boldsymbol{\theta}^{(k)})\epsilon_{kt}$. The proof relies on a set of preliminary results.

$$i)\ E\left\|\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0)\dot{\boldsymbol{\ell}}'_t(\boldsymbol{\theta}_0)\right\| < \infty,$$

$$ii)\ \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \text{vech}^0\left\{\boldsymbol{\eta}^*_t\left(\boldsymbol{\eta}^*_t\right)'\right\}}{\partial\boldsymbol{\theta}'} - \frac{\partial \text{vech}^0\left\{\tilde{\boldsymbol{\eta}}^*_t\left(\tilde{\boldsymbol{\eta}}^*_t\right)'\right\}}{\partial\boldsymbol{\theta}'}\right\| \to 0\ , \quad \text{in probability,}$$

$$iii)\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \text{vech}^0\left\{\boldsymbol{\eta}^*_t\left(\boldsymbol{\eta}^*_t\right)'\right\}}{\partial\boldsymbol{\theta}'}\right)_{\boldsymbol{\theta}_n} \to -\frac{1}{2}\boldsymbol{\Lambda}', \quad \text{a.s.} \quad \text{for any } \boldsymbol{\theta}_n \text{ between } \hat{\boldsymbol{\theta}}_n \text{ and } \boldsymbol{\theta}_0,$$

$$iv)\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0)\\ \text{vech}^0\left\{\boldsymbol{\eta}^*_t\left(\boldsymbol{\eta}^*_t\right)' - \boldsymbol{R}\right\}\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}\boldsymbol{J}^* & \boldsymbol{L}\\ \boldsymbol{L}' & \boldsymbol{\Gamma}\end{pmatrix}\right),$$

Point *i)* follows from the arguments given to prove i) in the proof of the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ (Theorem 3.1). Point ii) is equivalent to

$$\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\left(\eta^*_{kt}\eta^*_{\ell t}\right)}{\partial\boldsymbol{\theta}'}(\boldsymbol{\theta}) - \frac{\partial\left(\tilde{\eta}^*_{kt}\tilde{\eta}^*_{\ell t}\right)}{\partial\boldsymbol{\theta}'}(\boldsymbol{\theta})\right\| \to 0\ , \quad \text{in probability.}$$

In view of

$$\frac{\partial}{\partial\boldsymbol{\theta}'}\left\{\eta^*_{kt}\eta^*_{\ell t}\right\}(\boldsymbol{\theta}) = -\frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})}\left(\frac{1}{\sigma_{kt}}\frac{\partial\sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}'} + \frac{1}{\sigma_{\ell t}}\frac{\partial\sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})}{\partial\boldsymbol{\theta}'}\right),$$

and the same equality for $\partial\left\{\tilde{\eta}^*_{kt}\tilde{\eta}^*_{\ell t}\right\}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$, with σ_{kt} and $\sigma_{\ell t}$ replaced by $\tilde{\sigma}_{kt}$ and $\tilde{\sigma}_{\ell t}$, the conclusion follows by the arguments used to establish ii) in the proof of the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$.

Now we turn to iii). Note that

$$\frac{\partial}{\partial\boldsymbol{\theta}'}\left\{\eta^*_{kt}\eta^*_{\ell t}\right\}(\boldsymbol{\theta}_0) = -\eta^*_{kt}\eta^*_{\ell t}\left(\frac{1}{\sigma_{kt}}\frac{\partial\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}'} + \frac{1}{\sigma_{\ell t}}\frac{\partial\sigma_{\ell t}(\boldsymbol{\theta}_0^{(\ell)})}{\partial\boldsymbol{\theta}'}\right).$$

Thus, letting $d_{(k)} = \sum_{i=k+1}^{m} d_i$ and ${}_{(k)}d = \sum_{i=1}^{k-1} d_i$, with obvious conventions when $k = 1$ or $k = m$,

$$E\left(\frac{\partial}{\partial\boldsymbol{\theta}'}\left\{\eta^*_{kt}\eta^*_{\ell t}\right\}(\boldsymbol{\theta}_0)\right) = -\frac{1}{2}R_{k\ell}[(\mathbf{0}_{1\times{}_{(k)}d}\ \boldsymbol{\Omega}'_k\ \mathbf{0}_{1\times d_{(k)}}) + (\mathbf{0}_{1\times{}_{(\ell)}d}\ \boldsymbol{\Omega}'_\ell\ \mathbf{0}_{1\times d_{(\ell)}})]$$

Therefore, we have

$$E\left(\frac{\partial}{\partial\boldsymbol{\theta}'}\left(\text{vech}^0\left\{\boldsymbol{\eta}^*_t\left(\boldsymbol{\eta}^*_t\right)'\right\}\right)_{\boldsymbol{\theta}_0}\right) = -\frac{1}{2}\begin{pmatrix}\boldsymbol{\Omega}'\mathbf{M}_{21}\\ \boldsymbol{\Omega}'\mathbf{M}_{31}\\ \vdots\\ \boldsymbol{\Omega}'\mathbf{M}_{m,m-1}\end{pmatrix} = -\frac{1}{2}\boldsymbol{\Lambda}'.$$

By the law of large numbers, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\}}{\partial \boldsymbol{\theta}'}\right)_{\boldsymbol{\theta}_0} \to -\frac{1}{2}\boldsymbol{\Lambda}', \quad \text{a.s.}$$

To complete the proof of iii), we will show that similarly to (B.2), for any $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that, almost surely,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\left(\frac{\partial \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\}}{\partial \boldsymbol{\theta}'}\right)_{\boldsymbol{\theta}} - \left(\frac{\partial \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\}}{\partial \boldsymbol{\theta}'}\right)_{\boldsymbol{\theta}_0}\right\| \le \varepsilon.$$

The latter convergence is equivalent to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\ \sup_{\boldsymbol{\theta}^{(\ell)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(\ell)})}\left\|\frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})}\frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\partial\sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}'}\right.$$
$$\left. -\frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}_0^{(\ell)})}\frac{1}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\partial\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}'}\right\| \le \varepsilon, \quad a.s.$$

for any $k,\ell = 1,\dots,m$. By the arguments used to prove iii) in the proof of the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, we have

$$E\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\ \sup_{\boldsymbol{\theta}^{(\ell)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(\ell)})}\left\|\frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})}\frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\partial\sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}'}\right\| < \infty,$$

from which (B.5) follows. Thus, iii) is established.

It remains to show iv). We note that

$$\boldsymbol{Z}_t := \begin{pmatrix} \dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0) \\ \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)' - \boldsymbol{R}\right\} \end{pmatrix} = \begin{pmatrix} F(\boldsymbol{\eta}_t^*)\boldsymbol{d}_t \\ \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)' - \boldsymbol{R}\right\} \end{pmatrix}$$

is measurable with respect to the σ-field $\mathcal{F}_t$ generated by $\{\boldsymbol{\eta}_u^*, u \le t\}$. We have, using the independence of the sequence $(\boldsymbol{\eta}_t^*)$ under (2.7),

$$\begin{aligned} \mathrm{Var}(F(\boldsymbol{\eta}_t^*)\boldsymbol{d}_t) &= E\{F(\boldsymbol{\eta}_t^*)E(\boldsymbol{d}_t\boldsymbol{d}_t')F(\boldsymbol{\eta}_t^*)\} = \boldsymbol{J}^*, \\ \mathrm{Cov}\left[F(\boldsymbol{\eta}_t^*)\boldsymbol{d}_t, \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\}\right] &= E\left\{F(\boldsymbol{\eta}_t^*)\boldsymbol{\Omega}\left[\mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\}\right]'\right\} = \boldsymbol{L}. \end{aligned}$$

Thus, $\forall\lambda \in \mathbb{R}^{d+m(m-1)/2}$, the sequence $\{\lambda'\boldsymbol{Z}_t, \mathcal{F}_t\}_t$ is an ergodic, stationary and square integrable martingale difference. The conclusion follows from the central limit theorem of Billingsley (1961).

We are now in a position to complete the proof of Theorem 4.2. Since $\hat{\boldsymbol{\theta}}_n^{(k)}$ converges to $\boldsymbol{\theta}_0^{(k)}$, which stands in the interior of the parameter space by **A7**, the derivative of the criterion $\tilde{Q}_n^{(k)}$ is equal to zero at $\hat{\boldsymbol{\theta}}_n^{(k)}$. In view of point ii), we thus have by a Taylor expansion of $Q_n^{(k)}$ at $\boldsymbol{\theta}_0^{(k)}$,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)}-\boldsymbol{\theta}_0^{(k)}\right) \overset{o_P(1)}{=} -\left(\frac{1}{n}\sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_{ij}^{*(k)})}{\partial\theta_i^{(k)}\partial\theta_j^{(k)}}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial}{\partial\boldsymbol{\theta}^{(k)}}\ell_{kt}(\boldsymbol{\theta}_0^{(k)})$$

where the $\boldsymbol{\theta}_{ij}^{*(k)}$'s are between $\hat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$. Thus we have, using iii) and iv),

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right) \overset{o_P(1)}{=} -\boldsymbol{J}_0^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n \dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0).$$

Another Taylor expansion around $\boldsymbol{\theta}_0$ yields,

$$\begin{aligned}&\sqrt{n}(\hat{\boldsymbol{\rho}}_n-\boldsymbol{\rho}_0)\\ =\;& \frac{1}{\sqrt{n}}\sum_{t=1}^n \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'-\boldsymbol{R}\right\}+\frac{1}{n}\sum_{t=1}^n\frac{\partial}{\partial\boldsymbol{\theta}'}\left(\mathrm{vech}^0\left\{\tilde{\boldsymbol{\eta}}_t^*\left(\tilde{\boldsymbol{\eta}}_t^*\right)'\right\}\right)_{\tilde{\boldsymbol{\theta}}_n}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right),\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}_n$ is between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, and

$$\tilde{\boldsymbol{\eta}}_t^*=\tilde{\boldsymbol{\eta}}_t^*(\boldsymbol{\theta})=\tilde{\boldsymbol{D}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t \quad\text{and}\quad \tilde{\boldsymbol{D}}_t(\boldsymbol{\theta})=\mathrm{diag}\{\tilde{\sigma}_{1t}(\boldsymbol{\theta}^{(1)}),\dots,\tilde{\sigma}_{mt}(\boldsymbol{\theta}^{(m)})\}.$$

It follows that, using v) and vi), denoting by $\boldsymbol{I}$ the identity matrix of size $m(m-1)/2$ and by $\boldsymbol{0}$ is null matrix of size $d\times m(m-1)/2$,

$$\begin{pmatrix}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\\ \sqrt{n}(\hat{\boldsymbol{\rho}}_n-\boldsymbol{\rho}_0)\end{pmatrix} \overset{o_P(1)}{=} \begin{pmatrix}-\boldsymbol{J}_0^{-1} & \boldsymbol{0}\\ \frac{1}{2}\boldsymbol{\Lambda}'\boldsymbol{J}_0^{-1} & \boldsymbol{I}\end{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{Z}_t.$$

The asymptotic distribution of Theorem 4.2 thus follows from vii).

It remains to establish that $\boldsymbol{\Sigma}$ is non singular. By (B.6), it suffices to show that $\mathrm{Var}(\boldsymbol{Z}_t)$ is nonsingular. We will show that for any $\boldsymbol{x}=(\boldsymbol{x}_i)\in\mathbb{R}^d$, where $\boldsymbol{x}_i\in\mathbb{R}^{d_i}$, for any $\boldsymbol{y}=(y_{k\ell})\in\mathbb{R}^{m(m-1)/2}$ and any $c\in\mathbb{R}$,

$$\boldsymbol{x}'\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0)+\boldsymbol{y}'\mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'-\boldsymbol{R}\right\}=c,\ a.s. \quad\Rightarrow\quad \boldsymbol{x}=0 \text{ and } \boldsymbol{y}=0.$$

Assume that the left-hand side of (B.7) holds. Then we have

$$\sum_{i=1}^m\left(1-\eta_{it}^{*2}\right)z_{i,t-1}+\sum_{k\neq\ell}y_{k\ell}\left(\eta_{kt}^*\eta_{\ell t}^*-R_{k\ell}\right)=c$$

where $z_{i,t-1}=\frac{1}{\sigma_{it}^2}\boldsymbol{x}_i'\frac{\partial\sigma_{it}^2(\boldsymbol{\theta}_0^{(i)})}{\partial\boldsymbol{\theta}^{(i)}}$, that is,

$$(\boldsymbol{\eta}_t^*)'\boldsymbol{B}_{t-1}\boldsymbol{\eta}_t^*=c_{t-1}$$

for some symmetric matrix $\boldsymbol{B}_{t-1}$ and some number c_{t-1} belonging to the past. Thus $\boldsymbol{\eta}_t'\boldsymbol{R}^{1/2}\boldsymbol{B}_{t-1}\boldsymbol{R}^{1/2}\boldsymbol{\eta}_t = c_{t-1}$ from which it follows that, for $i = 1, \ldots, m$,

$$\mathrm{Cov}(\boldsymbol{\eta}_t'\boldsymbol{R}^{1/2}\boldsymbol{B}_{t-1}\boldsymbol{R}^{1/2}\boldsymbol{\eta}_t, \eta_{it}^2) = 0.$$

In view of Assumption **A13**, we deduce that $\boldsymbol{R}^{1/2}\boldsymbol{B}_{t-1}\boldsymbol{R}^{1/2}$ has a null diagonal. By replacing η_{it}^2 by $\eta_{kt}\eta_{\ell t}$ in the previous covariance, we similarly deduce that $\boldsymbol{R}^{1/2}\boldsymbol{B}_{t-1}\boldsymbol{R}^{1/2}$ or, equivalently, $\boldsymbol{B}_{t-1} = 0$. By noting that the diagonal terms of $\boldsymbol{B}_{t-1}$ are the $z_{i,t-1}$, we deduce by **A12** that $\boldsymbol{x} = 0$. It is then straightforward to show that $\boldsymbol{y} = 0$ and the proof is complete. □

B.7. Proof of Theorem 4.3

Let

$$O_n(\boldsymbol{\vartheta}) = \frac{1}{n}\log\frac{L_n(\boldsymbol{\vartheta})}{L_n(\boldsymbol{\vartheta}_0)}.$$

Following the lines of proof of Lemma 2 in Francq, Roussignol and Zakoïan (2001), we have

$$\lim_{n\to\infty}\frac{1}{n}\log L_n(\boldsymbol{\vartheta}) = \lim_{n\to\infty}\frac{1}{n}\log\left\|\prod_{i=2}^{n}\mathbb{P}(f_{\boldsymbol{\eta}_i^*})\right\| = E\log g_{\boldsymbol{\vartheta}}(\boldsymbol{\eta}_t^* \mid \boldsymbol{\eta}_{t-1}^*, \boldsymbol{\eta}_{t-2}^*, \ldots)$$

where $g_{\boldsymbol{\vartheta}_0}(\cdot \mid \boldsymbol{\eta}_{t-1}^*, \boldsymbol{\eta}_{t-2}^*, \ldots)$ denotes the density of $\boldsymbol{\eta}_t^*$ given the σ-field generated by $\boldsymbol{\eta}_{t-1}^*, \boldsymbol{\eta}_{t-2}^*, \ldots$ By Jensen's inequality, for $\boldsymbol{\vartheta} \in \Theta^*$ we thus have

$$\lim_{n\to\infty} O_n(\boldsymbol{\vartheta}) = E\log\frac{g_{\boldsymbol{\vartheta}}(\boldsymbol{\eta}_t^* \mid \boldsymbol{\eta}_{t-1}^*, \boldsymbol{\eta}_{t-2}^*, \ldots)}{g_{\boldsymbol{\vartheta}_0}(\boldsymbol{\eta}_t^* \mid \boldsymbol{\eta}_{t-1}^*, \boldsymbol{\eta}_{t-2}^*, \ldots)} \le \log E\frac{g_{\boldsymbol{\vartheta}}(\boldsymbol{\eta}_t^* \mid \boldsymbol{\eta}_{t-1}^*, \boldsymbol{\eta}_{t-2}^*, \ldots)}{g_{\boldsymbol{\vartheta}_0}(\boldsymbol{\eta}_t^* \mid \boldsymbol{\eta}_{t-1}^*, \boldsymbol{\eta}_{t-2}^*, \ldots)} = 0,$$

with equality iff $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$, in view of the identifiability constraints on Θ^*. Using the compactness of Θ^*, the conclusion follows by standard arguments. □

B.8. Proof of Theorem 5.1

Before proving Theorem 5.1 we will establish two lemmas. The first one shows that $\hat{\boldsymbol{\theta}}_n^{(k)}$ is a consistent estimator of $\boldsymbol{\theta}_0^{(k)}$.

LEMMA B.1. *Let the assumptions of Theorem 5.1 be satisfied. Then*

$$\hat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}, \quad a.s. \quad as\ n \to \infty.$$

Proof: It consists in verifying the conditions required in Theorem 3.1 for the convergence in (3.1).

The existence of a (unique) ergodic, non anticipative, strictly and second-order stationary solution $(\boldsymbol{\epsilon}_t)$ of Model (5.1), under the conditions given in the corollary, follows from Boussama , Fuchs and Stelzer (2011), Theorem 2.4. Thus **A1** holds with $s = 2$.

Recall that $\theta_5^{(k)} \in (0, 1)$ for all $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$. Straightforward calculation shows that

$$\begin{aligned} & |\sigma^2_{kt}(\boldsymbol{\theta}^{(k)}) - \sigma^2_{kt}(\boldsymbol{\theta}_0^{(k)})| \\ \leq \;& K\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}\| \sum_{i\geq 0} \left(\{\theta_{05}^{(k)}\}^i + \{\theta_5^{(k)}\}^i \right) (\epsilon^2_{1,t-i-1} + |\epsilon_{1,t-1}\epsilon_{2,t-1}| + \epsilon^2_{2,t-i-1}). \end{aligned}$$

It follows, using the fact that $\boldsymbol{\epsilon}_t$ belongs to L^2, that **A2** is satisfied. We similarly show that **A3** holds true, and **A4** is satisfied by definition of $\boldsymbol{\Theta}^{(k)}$.

Now we turn to **A5**. Suppose $\sigma_t(\boldsymbol{\theta}_0^{(k)}) = \sigma_t(\boldsymbol{\theta}^{(k)})$, that is

$$\begin{aligned} & \theta_{01}^{(k)} + \theta_{02}^{(k)}\epsilon^2_{1,t-1} + \theta_{03}^{(k)}\epsilon_{1,t-1}\epsilon_{2,t-1} + \theta_{04}^{(k)}\epsilon^2_{2,t-1} + \theta_{05}^{(k)}\sigma^2_{t-1} \\ = \;& \theta_1^{(k)} + \theta_2^{(k)}\epsilon^2_{1,t-1} + \theta_3^{(k)}\epsilon_{1,t-1}\epsilon_{2,t-1} + \theta_4^{(k)}\epsilon^2_{2,t-1} + \theta_5^{(k)}\sigma^2_{t-1}. \end{aligned}$$

Then there exists some non zero variables $a_{t-2}, b_{t-2}, c_{t-2}, d_{t-2}$ belonging to the past of $\boldsymbol{\eta}_{t-1}$ such that

$$a_{t-2} + b_{t-2}\eta^2_{1,t-1} + c_{t-2}\eta_{1,t-1}\eta_{2,t-1} + d_{t-2}\eta^2_{2,t-1} = 0.$$

Therefore, the distribution of $\boldsymbol{\eta}_t$ conditional to the past is degenerate. Since $\boldsymbol{\eta}_t$ is independent from the past, this means that the unconditional distribution of $\boldsymbol{\eta}_t$ is degenerate, in contradiction with the existence of a density around zero. Thus $a_{t-2} = b_{t-2} = c_{t-2} = d_{t-2} = 0$, from which we deduce that $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}_0^{(k)}$. Therefore, **A5** is verified. □

Now we turn to the asymptotic distribution. Assumption **A7** being in failure, we cannot use Theorem 3.2 to derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n^{(k)}$. It will be more convenient to work with a reparameterization. Consider the one-to-one transformation defined by $\boldsymbol{\Theta}^{(k)} \mapsto \boldsymbol{\Psi}^{(k)} = H(\boldsymbol{\Theta}^{(k)}) : \boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)' \mapsto H(\boldsymbol{x}) = (x_1, x_2, 2\sqrt{x_2 x_4} - x_3, x_4, x_5)'$. Write $\boldsymbol{\psi} = H(\boldsymbol{\theta})$. The following lemma derives the asymptotic distribution of $\hat{\boldsymbol{\psi}}_n^{(k)} = H(\hat{\boldsymbol{\theta}}_n^{(k)})$. Let $\boldsymbol{\Lambda} = \mathbb{R}^2 \times (0, \infty) \times \mathbb{R}^2$.

Lemma B.2. *Let the assumptions of Theorem 5.1 be satisfied. Then*

$$\sqrt{n}(\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) \quad \xrightarrow{\mathcal{L}} \quad \boldsymbol{\lambda}^{\boldsymbol{\Lambda}} := \arg\inf_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \{\boldsymbol{\lambda} - \boldsymbol{Z}\}' \dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})' \{\boldsymbol{\lambda} - \boldsymbol{Z}\}$$

where $\boldsymbol{Z} \sim \mathcal{N}\left\{0, \dot{\boldsymbol{H}}_k' \boldsymbol{J}_{kk}^{-1} \boldsymbol{I}_{kk} \boldsymbol{J}_{kk}^{-1} \dot{\boldsymbol{H}}_k\right\}$, *with* $\dot{\boldsymbol{H}}_k' = \frac{\partial H}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0^{(k)})$.

Proof: Note that, because $H_0^{(k)}$ is satisfied for the BEKK-GARCH(1,1) model, the third component of $\boldsymbol{\psi}_0^{(k)}$ is equal to zero, the other ones being strictly positive. We follow the lines of proof of Theorem 2 in Francq and Zakoian (2007). First note that the matrix $\dot{\boldsymbol{H}}_k$ is well defined (because $\theta_{02}^{(k)}, \theta_{04}^{(k)} > 0$) and is non-singular. Note also that, $\boldsymbol{\Lambda}$ being a convex cone, $\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$ is uniquely determined.

Except **A7**, the assumptions of Theorem 3.2 are satisfied. For instance, the verification of **A12** is achieved by the same arguments as those used for **A5**. For brevity, we do not detail the verification of all the assumptions. It follows in particular that J_{kk} is non singular.

A Taylor expansion of $H(\hat{\boldsymbol{\theta}}_n^{(k)})$ around $\boldsymbol{\theta}_0^{(k)}$ yields,

$$\sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\} \overset{o_P(1)}{=} \dot{\boldsymbol{H}}_k' \sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}),$$

using the convergence established in Lemma B.1 and the continuity of $\partial H/\partial \boldsymbol{\theta}$ (the notation $a_n \overset{o_P(1)}{=} b_n$ stands for sequences (a_n) and (b_n) such that $a_n - b_n$ converges to zero in probability). Now let

$$\boldsymbol{Z}_n = -\dot{\boldsymbol{H}}_k' \boldsymbol{J}_{kk}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (1 - \eta_{kt}^{*2}) \boldsymbol{d}_{kt}.$$

Note that we do not have equality (up to $o_P(1)$ terms) between $\boldsymbol{Z}_n$ and the left-hand side of (B.9) because, under $H_0^{(k)}$, the third component of this vector is a nonnegative random variable. This is not the case of $\boldsymbol{Z}_n$ which, by Theorem 3.2, converges in distribution to $\boldsymbol{Z}$.

We will establish that

$$\sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\} \overset{o_P(1)}{=} \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$$

where $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} = \arg\inf_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} \left\{\boldsymbol{\lambda} - \boldsymbol{Z}_n\right\}' \dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})' \left\{\boldsymbol{\lambda} - \boldsymbol{Z}_n\right\}$. Note that $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$ can be interpreted as the orthogonal projection of $\boldsymbol{Z}_n$ on $\boldsymbol{\Lambda}$ for the inner product $< x, y >_{\dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})'} = x' \dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})' y$. We also introduce the orthogonal projection of $\boldsymbol{Z}_n$ on $\sqrt{n}(\boldsymbol{\Psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})$, defined by

$$\tilde{\boldsymbol{\psi}}_n^{(k)} = \arg \inf_{\boldsymbol{\psi}^{(k)} \in \boldsymbol{\Psi}^{(k)}} \|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})\|_{\dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})'}.$$

Because $\sqrt{n}(\boldsymbol{\Psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})$ increases to $\boldsymbol{\Lambda}$, it can be noted that the variables $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$ and $\sqrt{n}\left\{\tilde{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\}$ are equal for n sufficiently large.

A Taylor expansion of the quasi-likelihood function yields

$$\begin{aligned}
&\tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) - \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)}) \\
= \;& \frac{\partial \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)\prime}}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) + \frac{1}{2}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})' \left[\frac{\partial^2 \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}}\right](\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) + R_n(\boldsymbol{\theta}^{(k)}) \\
= \;& -\frac{1}{2n}\boldsymbol{Z}_n' \dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk}\sqrt{n}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) - \frac{1}{2n}\sqrt{n}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})' \boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k')^{-1}\boldsymbol{Z}_n \\
& + \frac{1}{2}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})' \boldsymbol{J}_{kk}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) + R_n(\boldsymbol{\theta}^{(k)}) + R_n^*(\boldsymbol{\theta}^{(k)}) \\
= \;& \frac{1}{2n}\|(\dot{\boldsymbol{H}}_k')^{-1}\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})\|^2_{\boldsymbol{J}_{kk}} - \frac{1}{2n}\boldsymbol{Z}_n' \dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'\boldsymbol{Z}_n \\
& + R_n(\boldsymbol{\theta}^{(k)}) + R_n^*(\boldsymbol{\theta}^{(k)}) \\
= \;& \frac{1}{2n}\|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})\|^2_{\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'} - \frac{1}{2n}\boldsymbol{Z}_n' \dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'\boldsymbol{Z}_n \\
& + R_n(\boldsymbol{\theta}^{(k)}) + R_n^*(\boldsymbol{\theta}^{(k)}).
\end{aligned}$$

Following the lines of proof of Theorem 2 in Francq and Zakoian (2007), it can be shown that

i) $\sqrt{n}(\tilde{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) = O_P(1)$,

ii) $\sqrt{n}(\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) = O_P(1)$,

iii) for any sequence $(\boldsymbol{\theta}_n)$ such that $\sqrt{n}(\boldsymbol{\theta}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}) = O_P(1)$,

$$R_n(\boldsymbol{\theta}_n^{(k)}) = o_P(n^{-1}), \quad R_n^*(\boldsymbol{\theta}_n^{(k)}) = o_P(n^{-1}),$$

iv) $\|\boldsymbol{Z}_n - \sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\}\|^2_{\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'} \overset{o_P(1)}{=} \|\boldsymbol{Z}_n - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}\|^2_{\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'}$,

v) $\sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\} \overset{o_P(1)}{=} \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$,

vi) $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} \overset{\mathcal{L}}{\to} \boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$.

We omit the proof of these steps, which relies on arguments already given. The proof of Lemma B.2 then follows from v) and vi). □

Now we complete the proof of Theorem 5.1. Note that, from Example 8.2 in Francq and Zakoïan (2010), the third component of $\boldsymbol{\lambda}$ is the positive part, Z_3^+ say, of the third component of $\boldsymbol{Z}$. It follows that, letting $\boldsymbol{e}_3 = (0,0,1,0,0)'$,

$$\boldsymbol{e}_3'\sqrt{n}(\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) = \boldsymbol{e}_3'\sqrt{n}\hat{\boldsymbol{\psi}}_n^{(k)} \overset{\mathcal{L}}{\to} \boldsymbol{e}_3'\boldsymbol{\lambda}^{\boldsymbol{\Lambda}} = Z_3^+, \qquad Z_3 \sim \mathcal{N}\left\{0, \boldsymbol{e}_3'\dot{\boldsymbol{H}}_k'\boldsymbol{J}_{kk}^{-1}\boldsymbol{I}_{kk}\boldsymbol{J}_{kk}^{-1}\dot{\boldsymbol{H}}_k\boldsymbol{e}_3\right\}.$$

Noting that $\boldsymbol{e}_3'\dot{\boldsymbol{H}}_k' = \left(0, \sqrt{\theta_{04}^{(k)}/\theta_{02}^{(k)}}, -1, \sqrt{\theta_{02}^{(k)}/\theta_{04}^{(k)}}, 0\right)$, the conclusion straightforwardly follows from the consistency of $\boldsymbol{X}_n$, $\hat{\boldsymbol{J}}_{kk}$ and $\hat{\boldsymbol{I}}_{kk}$ to $\boldsymbol{e}_3'\dot{\boldsymbol{H}}_k$, $\boldsymbol{J}_{kk}$ and $\boldsymbol{I}_{kk}$ respectively. □

References

Aielli, G.P. (2013) Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics* 31, 282–299.

Baum L. E. (1972) An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities* 3, 1–8.

Baum L. E., and T. Petrie (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics* 30, 1554–1563.

Bauwens, L., Hafner, C. and S. Laurent (2012) *Handbook of volatility models and their applications.* Wiley.

Bauwens, L., Laurent, S. and J.V.K. Rombouts (2006) Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–109.

Berkes, I. and L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

Berkes, I., Horváth, L. and P. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

Billingsley, P. (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31, 307–327.

Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics* 72, 498–505.

Boussama, F., Fuchs, F. and R. Stelzer (2011) Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models. *Stochastic Processes and their Applications* 121, 2331–2360.

Cappé, O., Moulines, E. and T. Rydén (2005) *Inference in hidden Markov models.* Springer.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.

Engle, R.F. (2002) Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.

Engle, R.F. and B. Kelly (2012) Dynamic equicorrelation. *Journal of Business & Economic Statistics* 30, 212–228.

Engle, R.F. and K. Kroner (1995) Multivariate simultaneous GARCH. *Econometric Theory* 11, 122–150.

Engle, R.F. and K. Sheppard (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, University of California San Diego, Discussion paper.

Engle, R.F., Ng, V.K. and M. Rotschild (1990) Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* 45, 213-238.

Engle, R.F., Shephard, N. and K. Sheppard (2008) Fitting vast dimensional time-varying covariance models. NYU Working Paper FIN-08-009.

Escanciano, J.C. (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.

Fermanian, J.D. and H. Malongo (2014) Asymptotic theory of DCC models. Unpublished document.

Francq, C. and M. Roussignol (1997) On white noises driven by hidden Markov chains. *Journal of Time Series Analysis* 18, 553–578.

Francq, C., Roussignol, M. and J-M. Zakoïan (2001) Conditional Heteroskedasticity driven by Hidden Markov Chains, *Journal of Time Series Analysis* 22, 197-220.

Francq, C. and J-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. and J-M. Zakoïan (2007) Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and Their Applications* 117, 1265–1284

Francq, C. and J-M. Zakoïan (2010) *GARCH models: structure, statistical inference and financial applications.* Chichester: John Wiley.

Francq, C. and J-M. Zakoïan (2012) QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* 28, 179–206.

Frühwirth-Schnatter, S. (2005) *Finite mixture and Markov-switching models.* Springer.

Hamadeh, T. and J-M. Zakoïan (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141, 488–507.

Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

He, C. and T. Teräsvirta (2004) An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory* 20, 904–926.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.

Lanne, M. and P. Saikkonen (2007) A multivariate generalized orthogonal factor GARCH model. *Journal of Business & Economic Statistics* 25, 61–75.

Ling, S. (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140, 849–873.

Pelletier, D. (2006) Regime switching for dynamic correlations. *Journal of Econometrics* 131, 445–473.

Silvennoinen, A. and T. Teräsvirta (2009) Multivariate GARCH models. *Handbook of Financial Time Series* T.G. Andersen, R.A. Davis, J-P. Kreiss and T. Mikosch, eds. New York: Springer.

Sucarrat, G., Grønneberg, S. and A. Escribano (2013) Estimation and inference in univariate and multivariate LOG-GARCH-X models when the conditional density is unknown. MPRA Discussion Paper No. 49344.

Tse, Y.K. and A. Tsui (2002) A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351–362.

van der Weide, R. (2002) GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17, 549–564.

Yacowitz, S.J. and J. Spragins (1968) On the identifiability of finite mixtures. *The Annals of Mathematical Statistics* 39, 209–214.